

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



INTERNATIONAL BUREAU OF PATENT COOPERATION
35, rue de la Harpe, CH-1015, Yvertois, Suisse
P.O. Box 6859, CH-1000, Lausanne, Suisse
Rue de la Loi 50, B-1049, Bruxelles, Belgique
Rue de la Loi 50, B-1049, Bruxelles, Belgique

(43) International Publication Date
6 September 2002 (06.09.2002)

PCT

(10) International Publication Number
WO 02/068666 A1

- (51) International Patent Classification: C12N 15/82
- (21) International Application Number: PCT/US01/43588
- (22) International Filing Date:
16 November 2001 (16.11.2001)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
60/249,608 17 November 2000 (17.11.2000) US
- (71) Applicant (for all designated States except US): GENENCOR INTERNATIONAL, INC. [US/US]; 925 Page Mill Road, Palo Alto, CA 94304 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): DUNN-COLEMAN, Nigel [US/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). LANGDON, Timothy [GB/GB]; Troedryhiw, Lon Tylwyd, Llanfarian, Aberystwyth SY23 4UJ (GB). MORRIS, Phillip [GB/GB]; 6 Trefaenor, Comins-coch, Aberystwyth SY23 3UB, Cerdigion, Wales (GB).
- (74) Agent: BOYD, Victoria, L.; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304-1013 (US).
- (81) Designated States (national): AE, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, EE (utility model), EE, ES, FI (utility model), FI, GB, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NR, SN, TD, TG).
- Published:
— with international search report
- (48) Date of publication of this corrected version:
6 November 2003
- (15) Information about Correction:
see PCT Gazette No. 45/2003 of 6 November 2003, Section II
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

BEST AVAILABLE COPY

WO 02/068666 A1

(54) Title: MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES ENCODING CELL WALL DEGRADING ENZYMES

(57) Abstract: Described herein are methods to enhance the production of more highly fermentable carbohydrates in plants, especially forage grasses. The invention provides for transgenic plants transformed with expression vectors containing a DNA sequence encoding ferulic acid esterase I from *Aspergillus*, preferably *A. Niger*. The expression vectors may optionally comprise a DNA sequence encoding xylanase from *Trichoderma*, preferably *T. reesei*. Expression of the enzyme(s) is targeted to specific cellular compartments, in specific cellular compartments, in specific tissues and under specific environmental conditions. Uses of this invention include, but are not limited to, forage with improved digestibility for livestock, and enhanced biomass conversion.

5

10 **MANIPULATION OF THE PHENOLIC ACID CONTENT AND DIGESTIBILITY
OF PLANT CELL WALLS BY TARGETED EXPRESSION OF GENES
ENCODING CELL WALL DEGRADING ENZYMES**

15

**Inventors: Nigel Dunn-Coleman
Timothy Langdon
Phillip Morse**

CROSS-REFERENCE TO RELATED APPLICATIONS

20 Pursuant to 35 U.S.C. §119(e), the present application claims benefit of
and priority to USSN 60/249,608, entitled "MANIPULATION OF THE PHENOLIC
ACID CONTENT AND DIGESTIBILITY OF FORAGE GRASS CELL WALLS BY
TARGETED EXPRESSION OF A FERULIC ACID ESTERASE GENE", filed
November 17, 2000, by Morris et al.

FIELD OF THE INVENTION

25 This invention relates to methods to enhance to availability of fermentable
carbohydrates.

BACKGROUND OF THE INVENTION

30 The present crisis in livestock agriculture has prompted a resurgence of
interest in grass-fed animals. However, while a high-forage diet may be desirable, it
does not currently satisfy the demands of modern animal production. For the animal
to make efficient use of the forage it consumes, the energy demands of the micro-
organisms in the rumen must be met and synchronized with the availability of plant
proteins. Otherwise this lack of synchrony will lead to (a) proteins and other
35 nutrients being poorly utilized in the rumen, (b) loss of nitrogen, in urine and feces
and therefore, the environment and (c) the need to feed excessive amounts of
protein concentrates as supplements to the ruminant diet.

Cellulose and hemicellulose in grass and maize tissues could meet the

energy requirements of the ruminant or provide new feed-stocks for industrial fermentation to ethanol. This potential is not currently realized because the cell walls are lignified and the cell wall polysaccharides highly cross-linked with phenolic residues and lignin, resulting in low rates of plant cell wall digestion in comparison to rates of protein breakdown in ruminants. This is a particular problem for the most important forages in Europe, the ryegrasses *Lolium perenne* and *L. multiflorum* as well as one of the major impediments to the wider use of better adapted species, such as *Festuca arundinacea*, as a forage crop. Increasing the digestibility index of grasses has therefore been a major breeding objective for several decades but progress has been slow due to difficulties in fixing natural variation in the synthetic varieties derived from these outbreeding species (Hayward, *et al.*, TAG 70:48 (1985)).

Removing labile phenolics by chemical treatment with alkali is known to increase the biodegradability and nutritional value of low-quality feed such as cereal straw, and is employed commercially for feed upgrading. Reducing phenolic cross-linking of cell wall carbohydrates is therefore a predictable way of improving the rate of digestion and digestibility of ryegrass. However chemical modification may have other disadvantages. Therefore, genetic modification would be a preferable method of changing the cell wall chemistry of highly digestible varieties. Many in the field are pursuing this approach. An alternative, however, is to use genetic modification to reduce the levels of phenolic acids in the cell walls available for crosslinking either by directly disrupting ester bonds linking phenolics and lignins to cell wall polysaccharides or by preventing excessive ferulation of cell wall carbohydrates prior to their incorporation into the cell wall.

This invention meets this and other needs by using targeted or inducible expression of cell wall degrading enzymes in plants.

SUMMARY OF THE INVENTION

Provided herein are methods for enhancing the availability of fermentable carbohydrates. In one aspect, there is provided an expression cassette
5 comprising a DNA sequence encoding at least one cell wall degrading enzyme. The DNA sequence encoding at least one cell wall degrading enzyme may be operatively linked to a promoter sequence. The promoter may be constitutive or inducible. The expression cassette may further comprise a targeting sequence.

In one embodiment, the cell wall degrading enzyme is selected from the
10 group consisting of ferulic acid esterase, xylanase, xylosidase, cellulase, endoglucanase, and cellbiohydrolase. In a preferred embodiment cell wall degrading enzyme is derived from a fungal source. In a more preferred embodiment, the fungal ferulic acid esterase is an *Aspergillus* ferulic acid esterase, preferably *A. niger*. In another embodiment the xylanase is derived
15 from *Trichoderma*, preferably *T. reesei*.

In another aspect of the invention, there is provided a plant transformed with the expression cassette comprising a DNA sequence encoding at least one cell degrading enzyme. The plant may be selected from the group consisting of Festuca, Lolium, Avena and Zea. In a preferred embodiment the plant is a
20 forage grass. In another embodiment, the plant is maize.

Further provided herein is a method of controlling the level of phenolic acids in plant cell walls of a transgenic plant. The method, in one embodiment, comprises introducing to a plant cell an expression cassette comprising a DNA sequence encoding at least one cell wall degrading enzyme, preferably a ferulic
25 acid esterase.

Other objects, features and advantages of the present invention will become apparent from the following detailed description. It should be understood, however, that the detailed description and specific examples, while indicating preferred
30 embodiments of the invention, are given by way of illustration only, since various changes and modifications within the scope and spirit of the invention will become apparent to one skilled in the art from this detailed description.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 illustrates a restriction map of a DNA fragment containing the gene encoding the 38kd ferulic acid esterase.

Figures 2 A-E illustrate the complete DNA (SEQ. ID NO:__), with
5 highlighting to point out the signal sequence, intron and various restriction endonuclease sites, and amino acid sequence (SEQ. ID. NO:__) corresponding to the gene encoding the 38 kD ferulic acid esterase isolated from *Aspergillus niger*.

Figure 3 illustrates the DNA sequence of the gene encoding the 38 kD
10 esterase (SEQ. ID. NO:__).

Figure 4 illustrates the construction of the Intronless ferulic acid esterase isolated from *Aspergillus niger*.

Figure 5 illustrates that the overlapping of PCR products made with primers FAE-I5
FAE-I3 creates two possible uninterrupted reading frames – the top in the figure below is
15 functional (highlighted serine is at active site), the bottom is inactivated.

Figure 6 illustrates the possible vector constructions useful in the present invention. Various combinations are possible. Although and FAE gene is depicted another cell wall degrading enzyme may be used alone (i.e., instead of) or in conjunction with the FAE gene. Amp = ampicillin resistance gene.

20 Figure 7 illustrates pCOR105.

Figure 8 illustrates a generic ALE-TER vector.

Figure 9 illustrates the KDEL-COOH ER retention sequences.

Figure 10 illustrates the FAE-LINKER-FRAMESHIFT structure and sequence.

25 Figure 11 illustrates plant transformation cassettes.

Figure 12 is a table of the vectors used herein.

Figure 13 depicts the barley aleurain vacuolar and apoplast signal sequence.

Figure 14 illustrates the rat sialyl transferase structure and sequence.

30 Figure 15 illustrates the potato protease inhibitor II (PPI) motif structure and sequence.

Figure 16 illustrates the targeted expression of gfp to different cell compartment. Also shown are schematics of the vectors used.

Figure 17 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under ER and APO targeting sequences.

5 Figure 18 illustrates the FAE activity in transgenic *Festuca arundinacea* leaves of different ages under Vac targeting sequence.

Figure 19 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves of different ages.

10 Figure 20 illustrates the FAE activity in transgenic *Lolium mutflorum* leaves under Vac, ER and APO targeting sequences.

Figure 21 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under Vac targeting sequence.

15 Figure 22 illustrates the levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under APO and ER targeting sequence.

Figure 23 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Festuca arundinacea* plants expressing FAE under an actin promoter.

20 Figure 24 illustrates the *in vitro* dry matter digestibility of leaf tissue of mature *Lolium mutflorum* plants expressing FAE under an actin promoter.

Figure 25 illustrates the rate of fermentation and cumulative gas production in *Festuca arundinacea* cells.

Figure 26 illustrates the *in vitro* fermentation of *Festuca arundinacea* cell walls from cell cultures expressing recombinant FAE1.

25 Figure 27 illustrates the Time to maximum rate digestion for *Festuca arundinacea* cells.

Figure 28 illustrates the total gas production in *Festuca arundinacea* cells.

Figure 29 illustrates the kinetics of FAE activity by ferulic acid release from cell wall under self digestion in *Festuca arundinacea* and stimulation by xylanase.

30 Figure 30 illustrates the beta-glucoronidase activity under the *Lolium* See1 senescence promoter in leaves of transgenic plants of *Lolium mutflorum*.

Figure 31 illustrates the release of monomeric and dimeric HCAs on self digestion of leaves of vacuolar targeted FAE expressing plants.

Figure 32 is a schematic of the pTP10-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

5 Figure 33 is a schematic of the pUA4-4 vector. Also shown is the 5345 bp nucleotide sequence of the vector.

Figure 34 is a schematic of the pTU4 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

10 Figure 35 is a schematic of the pTT5.14 vector. Also shown is the 5395 bp nucleotide sequence of the vector.

Figure 36 is a schematic of the pTP8-5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 37 is a schematic of the pTP5-1 vector. Also shown is the 5277 bp nucleotide sequence of the vector.

15 Figure 38 is a schematic of the pTP4a2 vector. Also shown is the 5327 bp nucleotide sequence of the vector.

Figure 39 is a schematic of the pTP3-1 vector. Also shown is the 5338 bp nucleotide sequence of the vector.

20 Figure 40 is a schematic of the pTU5 vector. Also shown is the 5337 bp nucleotide sequence of the vector.

Figure 41 is a schematic of the pGT6 vector. Also shown is the 4773 bp nucleotide sequence of the vector.

Figure 42 is a schematic of the pJQ5 vector. Also shown is the 5034 bp nucleotide sequence of the vector.

25 Figure 43 is a schematic of the pJO6.1 vector. Also shown is the 4950 bp nucleotide sequence of the vector.

Figure 44 is a schematic of the pJQ4 vector. Also shown is the 4974 bp nucleotide sequence of the vector.

30 Figure 45 is a schematic of the pPQ10.1 vector. Also shown is the 5164 bp nucleotide sequence of the vector.

Figure 46 is a schematic of the pJQ3 vector. Also shown is the 4965 bp nucleotide sequence of the vector.

Figure 47 is a schematic of the pUG4 vector. Also shown is the 5295 bp nucleotide sequence of the vector.

Figure 48 is a schematic of the pUB8.11 vector. Also shown is the 5001 bp nucleotide sequence of the vector.

5 Figure 49 is a schematic of the pTP11-1 vector. Also shown is the 5387 bp nucleotide sequence of the vector.

Figure 50 illustrates the actin promoter and its corresponding nucleotide sequence.

10 Figure 51 illustrates the Aleurain-NPIR delete structure. The corresponding nucleotide sequences are also shown.

Figure 52 illustrates the SEE1 (senescence enhanced) promoter sequence.

Figure 53 illustrates the SEE1 (senescence enhanced) promoter sequence plus the vacuolar aleurain signal/NPIR sequence.

15

DETAILED DESCRIPTION OF THE INVENTION

The invention will now be described in detail by way of reference only using the following definitions and examples. All patents and publications, including all sequences disclosed within such patents and publications, referred
20 to herein are expressly incorporated by reference.

Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, *et al.*, DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D Ed., John Wiley and Sons, New York
25 (1994), and Hale & Marham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with a general dictionary of many of the terms used in this invention. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are
30 described. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy

orientation, respectively. Practitioners are particularly directed to Sambrook *et al.*, 1989, and Ausubel FM *et al.*, 1993, for definitions and terms of the art. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary.

5 The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole.

Definitions

10 It should be noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to a composition containing "a compound" includes a mixture of two or more compounds. It should also be noted that the term "or" is generally employed in its sense
15 including "and/or" unless the content clearly dictates otherwise.

 "Conservatively modified variants" applies to both amino acid sequences and polynucleotides. With respect to particular polynucleotides, conservatively modified variants refers to those polynucleotides that encode identical or essentially identical amino acid sequences, or where the polynucleotide does not
20 encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical polynucleotides encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the
25 corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every polynucleotide herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a polynucleotide (except AUG,
30 which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a polynucleotide which encodes a polypeptide is implicit in each described

sequence. For purposes of protein expression, there are "sub-optimal codons." These are codons that are not preferred by a particular genus or species. Altering these "sub-optimal codons" to "preferred codons" is a silent mutation in that the amino acid encoded by the codons is the same but one codon is
 5 preferentially expressed by the particular genus, *e.g.*, *Triticum spp.*

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a polynucleotide, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively
 10 modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

- 15 1) Alanine (A), Serine (S), Threonine (T);
 - 2) Aspartic acid (D), Glutamic acid (E);
 - 3) Asparagine (N), Glutamine (Q);
 - 4) Arginine (R), Lysine (K);
 - 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
 - 20 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
- (*see, e.g.*, Creighton, *Proteins* (1984)).

"Pyroglutamic acid" is the cyclized internal amide of L-glutamic acid

The phrase "controlling the level of phenolic acids" refers to the manipulation of phenolic acid expression in plants, particularly plant cell walls.
 25 The manipulation can be either positive; *e.g.*, increasing the levels of phenolic acids; negative, *e.g.*, decreasing the level of phenolic acids; or neutral, *e.g.*, changing the relative amounts of specific phenolic acids in the cell walls but keeping the total amount relatively the same. The timing of manipulation can be during plant growth or after plant growth, *e.g.*, after a plant has been cut or pulled
 30 from the ground or ingested. "Plant cell walls" refers to the cell walls of any cell of the plant.

The term "derived" means that a polynucleotide or protein is related to

another polynucleotide or protein. The relations can be one of homology, *e.g.*, nucleotides and proteins from certain species are homologous to similar polynucleotides and proteins of other species; analogy, *e.g.*, proteins perform the same function and therefore are related to each other regardless of organism of origin. The relationship can be a man-made one, *e.g.*, a protein (and a polynucleotide) can be derived from another protein by mutation; or chemical manipulation (peptidomimetics). Furthermore, a protein or a polynucleotide can be derived from an organism if, in the natural state, the protein or polynucleotide is found in one organism but recombinantly produced in another.

10 The term "exogenous polynucleotide" refers to a polynucleotide which is introduced into the plant by any means other than a sexual cross or sexual reproduction. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, *in planta* techniques, and the like. Such a plant containing the exogenous polynucleotide is referred to here as an R₁ generation transgenic plant. Transgenic plants which arise from sexual cross or by selfing are progeny of such a plant.

20 The term "isolated polynucleotide molecule" or "isolated protein" refers to a polynucleotide or protein which is essentially free of other cellular components with which it is associated in the natural state. It is preferably in a homogeneous state although it can be in either a dry or aqueous solution. Purity and homogeneity are typically determined using analytical chemistry techniques such as polyacrylamide gel electrophoresis or high performance liquid chromatography. A protein which is the predominant species present in a preparation is substantially purified. In particular, an isolated *FAE1* gene is separated from open reading frames which flank the gene and encode a protein other than *FAE1*.

25 A "*FAE1* encoding polynucleotide" is a nucleic acid sequence comprising (or consisting of) a coding region of an *FAE 1* gene or which encodes a *FAE1* polypeptide. *FAE1* polynucleotides can also be identified by their ability to hybridize under low stringency conditions (see below) to nucleic acid probes having a sequence of 8 to 300 bases, preferably a sequence of 80 to 100 bases

in the sequence disclosed in WO 98/14594.

The term "nucleic acid encoding," "nucleic acid sequence encoding" or "polynucleotide encoding" refers to a polynucleotide which directs the expression of a specific protein or peptide. The polynucleotides include both the DNA strand
5 sequence that is transcribed into RNA and the RNA sequence that is translated into protein. The polynucleotides include both full length polynucleotides as well as shorter sequences derived from the full length sequences. It is understood that a particular polynucleotide includes the degenerate codons of the native sequence or sequences which may be introduced to provide codon preference in
10 a specific host cell. The polynucleotide includes both the sense and antisense strands as either individual single strands or in the duplex form.

The term "operably linked" refers to functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates transcription of RNA corresponding to the second sequence.

15 The term "plasmid" refers to a circular double stranded DNA molecule which comprises the coding sequence of interest, regulatory elements, a selection marker and optionally an amplification marker. A plasmid can transform prokaryotic cells or transfect eukaryotic cells. An "expression cassette" means a portion of a plasmid (or the entire plasmid) containing the regulatory
20 elements desired for transcription, translation and/or expression and the coding region of a polynucleotide. A plasmid may contain one or more expression cassettes. If multiple expression cassettes are introduced into a plant, they can be introduced simultaneously or at different times. If simultaneous introduction is desired, the expression cassettes can be on one plasmid or more. Typically, an
25 expression cassette comprises a promoter, poly A+ tail, and signal sequences that target the expressed polypeptide to a specific region of a cell or to be secreted, if desired. Examples of signal sequences that "target expression" of ferulic acid esterase include sequences located upstream of the FAE coding sequence. The polynucleotide that encodes the signal sequence is found
30 preferably within the 100 nucleotides "upstream" (in the 5' direction) from the initiation codon (AUG). More preferably, the polynucleotide that encodes the signal sequence is found within the 50 nucleotides upstream from the initiation

codon. Many different cellular organelles are targeted by the signal sequences used in this invention. The organelles include, but are not limited to, vacuoles, Golgi apparatus, endoplasmic reticula, and apoplasts. In addition to upstream signal sequences, the expression cassette of this invention may include a
5 polynucleotide that encodes a signal sequence at the 3' end. These signal sequences include, but are not limited to stop codons and the KDEL sequence. In addition to KDEL, other similar sequences are contemplated by this invention, including but not limited to RDEL. In addition to a KDEL sequence, a signal sequence can include a linker to a KDEL sequence. A linker is an extension of
10 the reading frame of the encoding polynucleotide to the signal sequence. Preferably, the polynucleotide encoding the signal sequence is directly downstream from the coding sequence, more preferably less than 100 base pairs from the stop codon, more preferably less than 20 base pairs from the stop codon.

15 The term "polynucleotide," "polynucleotide" or "nucleic acid sequence" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. Unless specifically limited, the term encompasses polynucleotides containing known analogues of natural nucleotides which have similar binding properties as the reference polynucleotide
20 and are metabolized in a manner similar to naturally occurring nucleotides. Unless otherwise indicated, a particular *FAE1* polynucleotide of this invention also implicitly encompasses conservatively modified variants thereof (e.g. degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions
25 may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); and Cassol *et al.*, 1992; Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). The term polynucleotide is used interchangeably with
30 gene, cDNA, and mRNA encoded by a gene.

The term "polypeptide," "peptide," and "protein" are used interchangeably and refer to amino acids connected by peptide bonds. Polypeptides can be

entire proteins or portions thereof. For Example, a FAE1 polypeptide may refer to the entire FAE1 protein or fragments of the FAE1 protein. A "ferulic acid esterase with an altered glycosylation site" refers to a FAE protein wherein a mutation has changed the glycosylation pattern of the protein. Mutations that effect such changes are well known in the art and include, but are not limited to, amino acid substitutions, and mutations in the proteins of the Golgi apparatus and endoplasmic reticulum that effect glycosylation of proteins.

The term "promoter" refers to a polynucleotide that directs expression of a coding sequence. A promoter can be constitutive, *i.e.*, relatively independent of the stage of differentiation of the cell in which it is contained or it can be inducible, *i.e.*, induced by specific environmental factors, such as the length of the day, the temperature, *etc.* or a promoter can be tissue-specific, *i.e.*, directing the expression of the coding sequence in cells of a certain tissue type. A "senescence" promoter is an inducible promoter that causes transcription to be initiated upon a certain event relating to age of the organism. A "heat shock promoter" is an inducible promoter that causes transcription to be initiated upon a change in temperature. An example of a heat shock protein promoter is the Soybean Gmhsp promoter. In addition to these inducible promoters, one of skill will realize that other inducible promoters can be used. For example, a wound induced promoter, like LAP. See, US Patent No. 5,962,670.

The term "purified" denotes that a polynucleotide or protein gives rise to essentially one band in an electrophoretic gel. Particularly, it means that the polynucleotide or protein is at least 85% pure, more preferably at least 95% pure, and most preferably at least 99% pure.

The term "specifically hybridizes" refers to a nucleic acid probe that hybridizes, duplexes or binds to a particular target DNA or RNA sequence when the target sequences are present in a preparation of total cellular DNA or RNA. "Complementary" or "target" nucleic acid sequences refer to those nucleic acid sequences which selectively hybridize to a nucleic acid probe. Proper annealing conditions depend, for example, upon a probe's length, base composition, and the number of mismatches and their position on the probe, and must often be determined empirically. For discussions of nucleic acid probe design and

annealing conditions, see, for example, Sambrook *et al.*, MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) ("Sambrook") or CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, F. Ausubel *et al.*, ed. Greene Publishing and Wiley-Interscience, New York (1987) ("Ausubel").

- 5 The term "stringent conditions" in the context of polynucleotide hybridization experiments such as Southern and northern hybridizations refers to sequence dependent, binding and washing environments. An extensive guide to the hybridization of polynucleotides is found in Tijssen (1993) LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY—HYBRIDIZATION WITH
- 10 NUCLEIC ACID PROBES part I chapter 2 "overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York. Generally, highly stringent hybridization and wash conditions are selected to be about 5°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and
- 15 pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Very stringent conditions are selected to be equal to the T_m for a particular probe. An example of stringent hybridization conditions for hybridization of complementary polynucleotides which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin
- 20 with 1 mg of heparin at between 40 and 50°C, preferably 42°C, with the hybridization being carried out overnight. An example of highly stringent wash conditions is 0.15M NaCl at from 70 to 80°C with 72°C being preferable for about 15 minutes. An example of stringent wash conditions is a 0.2x SSC wash at about 60 to 70°C, preferably 65°C for 15 minutes (see, Sambrook, *supra* for a
- 25 description of SSC buffer). Often, a high stringency wash is preceded by a low stringency wash to remove background probe signal. An example medium stringency wash for a duplex of, e.g., more than 100 nucleotides, is 1x SSC at 40 to 50°C, preferably 45°C for 15 minutes. An example low stringency wash for a duplex of, e.g., more than 100 nucleotides, is 4-6x SSC at 35 to 45°C, with 40°C
- 30 being preferable, for 15 minutes. In general, a signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization. Polynucleotides which do

not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, *e.g.*, when a copy of a polynucleotide is created using the maximum codon degeneracy permitted by the genetic code.

5 The term "transgenic plant" refers to a plant into which exogenous polynucleotides have been introduced and their progeny. Typically, cells of a plant are transformed with the exogenous polynucleotide and a transgenic plant is regenerated from the transformed cells. The regenerated plant is then bred to produce a strain of transgenic plants.

10 "Xylanase" (EC 3.2.1.8) refers to a well described class of glycosyl hydrolases that hydrolyze xylan. Commercial applications of xylanase include the degradation and bleaching of wood pulp for paper making. Xylanase can also be added to animal feed to improve the digestibility of plant matter. Typically, commercial xylanase is derived from fungi. A preferred xylanase is derived from
15 *Trichoderma*.

Preferred Embodiments

Plant cell walls contain a range of alkali-labile ester-linked phenolic acids. In particular, grass cell walls are characterized by the presence of large amounts of esterified ferulic and *p*-coumaric acids (mainly in their *E* configurations), linked
20 to arabinoxylans at the C5 of arabinose. These are released as ferulated oligosaccharides (FAX and PAX) by cellulase treatment but *in vivo* provide a substrate for peroxidase-catalyzed cross-linking of cell wall polysaccharides and lignin. The high levels of these phenolic acids and their dimers have a dramatic influence on the mechanical properties, digestibility and rates of digestion of
25 grasses by ruminants.

Previous work has shown that ferulic acid is the predominant *p*-hydroxycinnamic acid esterified to grass polysaccharide but until recently the only ferulic acid dehydrodimer to have been isolated was 5,5'-diferulic acid. Recently new dehydrodiferulate dimers and cyclobutane-type dimer mixtures have been
30 isolated from plant cell walls (Waldron, *et al.*, *Phytochemical Analysis* 7:305 (1998)). As can be seen in Figure 1, these mixtures are present in large amounts in grass cells. Ether linked ferulic acid-coniferyl alcohol dimers, have also been isolated from cell walls (Jacquet, *et al.*, *Polyphenol Comm. Bordeaux*

pp451 (1996)) establishing for the first time that ferulate esters are oxidatively co-polymerized with lignin precursors which may anchor lignins to cell wall polysaccharides. The yield of these dimers in grass cells indicates that phenolic dehydrodimer cross-linking of cell wall polysaccharides is much more extensive
5 than was previously thought.

An enzyme system has been reported from parsley endomembranes that catalyses the ferulation of endogenous polysaccharide acceptors from feruloyl CoA, pointing to the ER/golgi as the site of polysaccharide esterification and the CoA ester as the physiological co-substrate (Meyer, *et al.*, *FEBS Lett.* 290:209
10 (1991)). Further evidence for this has been found in water-soluble extracellular polysaccharides excreted in large amounts into the medium by grass cell cultures. This material is highly esterified with ferulic and p-coumaric acid at levels similar to the cell walls of the cultured cells.

Feruloyl esterase activity has been detected in several fungal species
15 including, anaerobic gut fungi, yeasts, actinomycetes, and a few fiber-degrading ruminal bacteria, which enables them to de-esterify arabinoxylans and pectins.

Two ferulic acid esterases (FAE), distinguished on the basis of molecular weight and substrate specificity, have been isolated from *Aspergillus niger* and have been shown to quantitatively hydrolyze ferulic acid and release
20 dehydrodiferulate dimers from plant cell walls. Furthermore, FAE has been observed to act synergistically with xylanase to release ferulic acid from plant cell walls at a higher rate. Recently, a ferulic acid esterase (FAE) gene has been cloned from *Aspergillus niger* (Michelson, *et.al.* European Patent Application No. 9510370.1). The inventors have found the recombinant enzyme releases ferulic
25 acid and diferulate dimers from grass cell walls in a concentration dependent manner and that this enzyme is stable at 30°C pH 5.0 in the presence of substrate and has a half life of 61 h at 30°C in the presence of vacuolar extracts (pH 4.6) of grass cells. This gene was, therefore, a candidate for targeted and inducible expression of FAE in grasses (*e.g.*, *Lolium multiflorum*).

30 The present invention provides for methods of changing the cell wall structure of transgenic plants and therefore, making them more digestible. The method comprises introducing a ferulic acid esterase coding sequence into the

cells of a plant. Operably linked to the coding sequence is a promoter that can be either constitutive or inducible and signal sequences that serve to target expression of the coding sequence in the desired organelle in the desired cell of the plant. The signal sequences can be either or both N terminal or C terminal sequences.

Optionally, a second and/or third coding sequence is introduced into the plant. It is preferred that a fungal xylanase coding sequence be coexpressed with the FAE coding sequence.

This invention also provides for transgenic plants which contain FAE1 coding sequences, leading to more digestible grasses.

Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. Basic texts disclosing the general methods of use in this invention include Sambrook, *et al.*, MOLECULAR CLONING, A LABORATORY MANUAL, 2ND ED. (1989); Kriegler, GENE TRANSFER AND EXPRESSION: A LABORATORY MANUAL (1990); and Ausubel *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1994)).

A. Isolation of Polynucleotides

The isolation of the polynucleotides, *e.g.*, FAE1 and xylanase of the invention may be accomplished by a number of techniques. See, for example, copending US application 08/952,445 which describes the isolation of a FAE from *Aspergillus niger*; and copending US application 09/658,772 which describes the isolation of a xylanase from *T. reesei*.

For instance, oligonucleotide probes based on the sequences cited here can be used to identify the desired gene in a cDNA or genomic DNA library from a desired species. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, *e.g.*, using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be

packaged into the appropriate vector. To prepare a library of cDNA from a specific cell culture, e.g., *Aspergillus niger*, mRNA is isolated from the culture and a cDNA library containing the gene transcripts is prepared from the mRNA.

The cDNA or genomic library can then be screened using a probe based
5 upon the sequence of a known polynucleotide such as the polynucleotides cited here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. In addition to probes derived from known polynucleotides, degenerate probes may be used. Techniques for making and using degenerate probes are well known in the art
10 and can be found in Sambrook and Ausubel.

Alternatively, the polynucleotides of interest can be amplified from polynucleotide samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of the genes directly from mRNA, from cDNA, from genomic libraries
15 or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone polynucleotides that code for proteins to be expressed, to make polynucleotides to use as probes for detecting the presence of the desired mRNA in samples, for polynucleotide sequencing, or for other purposes.

Appropriate primers and probes for identifying ferulic acid esterase-specific genes, as well as xylanase sequences, from fungi and plant tissues are generated from comparisons of the sequences provided herein. For a general overview of PCR see PCR PROTOCOLS: A GUIDE TO METHODS AND APPLICATIONS, (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), Academic Press, San
20 Diego (1990). Reaction components are typically: 10 mM Tris-HCl, pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.001% gelatin, 200 μ M dATP, 200 μ M dCTP, 200 μ M dGTP, 200 μ M dTTP, 0.4 μ M primers, and 100 units per mL Taq polymerase. Program: 96°C for 3 min., 30 cycles of 96°C for 45 sec., 50°C for 60 sec., 72°C for 60 sec, followed by 72°C for 5 min.
25

Polynucleotides may also be synthesized by well-known techniques as described in the technical literature. See, e.g., Carruthers, *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams, *et al.*, *J. Am. Chem.*
30

Soc. 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

5 Suitable sources for the ferulic acid esterase used in this invention include but are not limited to, *Neurospora crassa*, *Aspergillus spp.* and specifically, *Aspergillus niger*. The xylanase used in this invention can be derived from any suitable source including, but not limited to, *Trichoderma reesei* and *Aspergillus spp.*

10

B. Preparation of Recombinant Vectors

To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of plant species are well known and described in the technical and scientific literature. See, for example, Weising, *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding the full length FAE1 protein, will preferably be combined with transcriptional and translational initiation and targeting regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant under the desired conditions.

Promoters can be identified by analyzing the 5' sequences of a desired gene. Sequences characteristic of promoter sequences can be used to identify the promoter. Sequences controlling eukaryotic gene expression have been extensively studied. For instance, promoter sequence elements include the TATA box consensus sequence (TATAAT), which is usually 20 to 30 base pairs upstream of the transcription start site. In most instances the TATA box is required for accurate transcription initiation. In plants, further upstream from the TATA box, at positions -80 to -100, there is typically a promoter element with a series of adenines surrounding the trinucleotide G (or T) N G. Messing, *et al.*, in GENETIC ENGINEERING IN PLANTS, pp. 221-227 (Kosage, Meredith and Hollaender, eds. (1983)).

A number of methods are known to those of skill in the art for identifying and characterizing promoter regions in plant genomic DNA (see, e.g., Jordano, *et al.*, *Plant Cell* 1:855-866 (1989); Bustos, *et al.*, *Plant Cell* 1:839-854 (1989); Green, *et al.*, *EMBO J.* 7:4035-4044 (1988); Meler, *et al.*, *Plant Cell* 3:309-316
5 (1991); and Zhang, *et al.*, *Plant Physiology* 110:1069-1079 (1996)).

In construction of recombinant expression cassettes of the invention, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions
10 and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, the actin and ubiquitin promoters and other transcription initiation regions from various plant genes known to those of skill. A particularly preferred constitutive
15 promoter is the rice actin promoter (see, McElroy, *Plant Cell*, 2:163 (1990)).

Alternatively, the plant promoter may direct expression of the polynucleotide of the invention in a specific tissue (tissue-specific promoters) or may be otherwise under more precise environmental control (inducible promoters). Examples of tissue-specific promoters under
20 developmental control include promoters that initiate transcription only in certain tissues, such as leaves, roots or seeds.

In one aspect of the instant invention, expression of FAE occurs after the the plant has been cut, removed from the ground or ingested. Thus an appropriate promoter would be a senescence promoter. For example, *BFN1* has
25 recently been shown to be a nuclease expressed in senescing leaves, Perez-Amador, *et al.*, *Plant Physiol.* 122:169 (2000). Similarly, SAG12, a cysteine protease is also found in senescing leaves (Noh & Amasino, *Plant Mol. Biol.* 41:181 (1999). In a preferred embodiment, the promoter from the *gem* gene of *Festuca pratensis* is used to direct expression of FAE in senescing leaves.

30 In another aspect, the FAE would be expressed upon ingestion by a foraging animal. Exemplary promoters for this aspect would include Soybean Gmhsp 17.5 promoter and the leucine aminopeptidase (LAP) promoter. The

GMhsp promoter is from a heat shock protein gene and initiates expression if the temperature of the environment is increased. In the laboratory, an increase of 15°C for 2 hours is the preferred heat shock. However, in non-laboratory conditions suitable increases in temperature will occur in silos and in the rumen of animals that have ingested the plants of this invention. The LAP promoter initiates the expression of the FAE gene upon wounding of the plant. Such wounding would occur after cutting the plant or after mastication by a foraging animal. Tissue specific promoters that could be used in this invention include promoters of genes that are differentially expressed in the leaves of grasses. An example of a leaf specific promoter is the *rbcS* promoter of tomato (*Proc. Nat'l Acad. Sci. USA* 84:7104 (1987)). This promoter normally regulates a gene determined to be important in photosynthesis.

For proper polypeptide expression, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural fungal gene, from a variety of other fungal or plant genes, or from T-DNA. These sequences are well known and readily available to those of skill in the art.

In addition to a promoter and poly A+ sequences, the preferred expression vectors of this invention also will contain signal sequences. These are polynucleotides found at the 5' and/or 3' ends of the coding region and serve to target expression of the gene to specific cellular organelles. These signal sequences can be both upstream or downstream of the coding region. Some preferred examples of upstream signal sequences include the barley aleurain sequence (Rogers, *Proc. Nat'l Acad. Sci. USA* 82:6512 (1985) which targets vacuoles and the *Aspergillus* apoplast signal. This signal sequence targets expression to the apoplast.

In addition to targeting expression to specific organelles, it may be desirable to retain the expressed FAE in the Golgi or endoplasmic reticulum. The well known ER retention signal, KDEL, can be added to the 3' end of the coding polynucleotide.

The vector comprising the expression cassettes (e.g., promoters and/or coding regions) of the invention will typically comprise a marker gene which

confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to hygromycin, kanamycin, G418, bleomycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

5

C. Production of Transgenic Plants

DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment or the constructs may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

See Dalton et al. (Co-transformed, diploid *Lolium perenne* (Perennial Ryegrass), *Lolium multiflorum* (Italian Ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. Plant Cell Reports (1999) 18(9), 721-726) for exemplary methods for culturing and transformation of grasses.

Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski, et al., *Embo J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm, et al., *Proc. Natl. Acad. Sci. USA* 82:5824 (1985).

Agrobacterium tumefaciens-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. See, for example Horsch, et al., *Science* 233:496-498 (1984), and Fraley, et al., *Proc. Natl. Acad. Sci. USA* 80:4803 (1983). US Patent 5,591,616 discloses *Agrobacterium* mediated transformation techniques in monocotyledons.

Ballistic transformation techniques are described in Klein, et al., *Nature*

327:70-73 (1987). In a preferred embodiment, a particle in-flow gun (PIG) is used to transform the plant cells of this invention.

Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant that possesses the transformed genotype and thus the desired phenotype such as improved digestibility. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans, *et al.*, PROTOPLASTS ISOLATION AND CULTURE, HANDBOOK OF PLANT CELL CULTURE, pp. 124-176, Macmillan Publishing Company, New York, 1983; and Binding, *REGENERATION OF PLANTS, PLANT PROTOPLASTS*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee, *et al.*, *Ann. Rev. of Plant Phys.* 38:467-486 (1987).

To determine the presence of or increase of FAE1 activity, an enzymatic assay can be used or an assay to measure increases and decreases in rates of fermentation. These assays are readily available in the literature and those of skill in the art can readily find them.

One of skill will recognize that other assays can be used to detect the presence or absence of FAE1. These assays include but are not limited to; immunoassays and electrophoretic detection assays (either with staining or western blotting).

The polynucleotides of the invention can be used to confer desired traits on essentially any plant. However, the main utility of this invention is in the improved digestibility of forage plants. Thus, it is envisioned the transgenic plants of this invention will include but not be limited to the following genera *Lolium*, *Festuca*, *Triticum*, *Avena*, and *Medicago*. The FAE1 genes of the invention are particularly useful in the production of transgenic plants in the genus *Lolium*.

One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be

introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

As mentioned above, the transgenic plants of this invention can be used as a foraging crop for animals, such as cattle, sheep, goats and horses. In addition, the methods of this invention can be used to transform any plant into which FAE expression is desired. For example, it is advantageous to break down cell walls during biomass conversion or during processing of plants for foodstuffs. This invention would help to achieve this goal more effectively and inexpensively.

10

The inventive methods herein may also be used to provide additional enzymes to enhance the availability of fermentable sugars in plants. Plant carbohydrates may be subject to further modification, either exogenously or endogenously, by the action of other enzymes. Such enzymes include, but are not limited to, endoglucanases, xylosidases and/or cellbiohydrolases. These enzymes may be provided either in an expression cassette provided for herein (i.e., endogenous) or applied to the plant cell walls (i.e., exogenous) to enhance the availability of mono- and/or di-saccharides.

Plants other than grasses may find a use in the present invention. For example, corn (or maize) is specifically contemplated to be useful. The grass Festuca is similar to maize in cell wall structure and therefore provides a good model of the ability to enhance fermentable carbohydrates in corn. Other useful plants contemplated for use in the present invention are Festuca, Lolium, Zea, Avena, Sorghum, Millet (tropical cereals), Miscanthus (a grass with potential for use as a biomass energy crop), Cenchrus, Dichanthium, Brachiaria and Paspalum (apomictic tropical range grasses) and Poa (Kentucky bluegrass).

Cell walls of forage grasses makes up 30-80% of forage dry matter representing a major source of energy for ruminants, but less than 50% of this fraction is digested by the animal. Conversion of low-value biomass to sugars and ethanol is also less than optimal due to the carbohydrate unavailability of the

feedstocks, including but not limited to bagasse, rice straw, corn stover and corn fiber.

5 Ferulic and other hydroxycinnamic acids are ester linked to arabinosyl residues in arabinoxylans, and play a key role in crosslinking xylans to lignin, resulting in less degradable cell walls. Ferulic acid esterase (FAE) can release both monomeric and dimeric ferulic acid (FA) from arabinoxylans making the cell wall more susceptible to further enzymatic attack. Transgenic plants have been produced expressing an FAE gene following microprojectile bombardment of cell cultures. Measurements of the level of FAE activity from different vectors
10 targeting FAE to the vacuole, ER and apoplast under constitutive or inducible (heat shock) promoters shows that at least for constitutive expression of vacuolar targeted FAE, the activity was highest in young leaves and increased along the leaf lamina. We also show that FAE expression results in release of monomeric and dimeric FA from cell walls on cell death and this was enhanced several fold
15 by the addition of xylanase. An effect of FAE expression on the monomeric and dimeric cell wall ester linked ferulate content in comparison to control (non-transformed) plants is seen. Generally, the lower the levels of monomers and, in particular, dimers of hydroxycinnamic acids in leaves, the higher the digestibility and/or availability of complex carbohydrates for conversion.

20 Senescence is the terminal phase in leaf development and occurs without growth or morphogenesis. Therefore the metabolism/physiology of this stage of the leaf's lifespan can be targeted directly for alteration with minimal detrimental impact on early development. Senescence follows leaf maturity and is associated with the expression of specific genes. These genes and their
25 controlling elements can be exploited to manipulate development, adaptation, productivity and quality traits in crop plants. There seems to be good conservation of senescence physiology across the range of higher plant species and thus these promoters are useful in the present invention.

30 The following preparations and examples are given to enable those skilled in the art to more clearly understand and practice the present invention. They should not be considered as limiting the scope and/or spirit of the invention, but

merely as being illustrative and representative thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); kg (kilograms); μ g (micrograms); L (liters); ml (milliliters); μ l (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); ° C. (degrees Centigrade); h (hours); min (minutes); sec (seconds); msec (milliseconds); Ci (Curies) mCi (millicuries); μ Ci (microCuries); TLC (thin layer chromatography); Et (ethyl), Me (methyl).

10

Example 1

Preparation of Enzyme Encoding DNA Sequences

A genomic clone for FAE1 (see Figures 1-3) was used as the starting point for the preparation of an intronless FAE1 encoding DNA sequence. The sequence for the genomic clone is given in Figures 2 and 3. Separate fragments for both FAE exons were recovered by PCR from a 5.5kb EcoRI fragment of the genomic clone in pLITMUS28, and 'cDNA' created by overlapping PCR. See Figure 4.

Two 5' primers were used. FAE-S5 which amplifies the entire reading frame (including the Aspergillus signal), and FAE-N5 which amplifies only the mature protein (i.e. has no signal). A number of codons are optimised (underlined in primer sequences below). The overlap product may be derived from either FAE-I5 (wild type) or FAE-I3 (conserved Ser changed to Ala) primers, allowing production of enzymatically inactive protein to check toxicity. As shown in Figure 5, overlapping of PCR products made with FAE-I5 and FAE-I3 creates two possible uninterrupted reading frames. If the complement to FAE-I5 serves as the template when recombined then the encoded protein retains the serine moiety and the esterase is functional (highlighted serine is at active site). If the FAE-I3 primer serves as the template the serine is replaced with an alanine and the esterase is inactivated (highlighted alanine in bottom amino acid sequence given in Figure 5).

25
30

Where possible, codon usage has been optimised in constructed reading frames (codon choice based on published barley preferences).

FAE-I5 (SEQ ID NO: ____)

5 GGCGCCGAGGGAGTGGCCGGTCACGGTCAGCGCGTAGTCC 40-mer

FAE-I3 (SEQ ID NO: ____)

CCGGCCACGCCCTCGGCGCCTCCCTGGCGGCACTC 35-mer

FAE-N5 (SEQ ID NO: ____)

CTAAAGCTTACCATGGCGGCCGCCTCCACGCAGGGCATCTCCGA 44-mer

10 **FAE-S5 (SEQ ID NO: ____)**

CTAAAGCTTAACATGAAGCAGTTCTCCGCCAA 32-mer

FAE-3 (SEQ ID NO: ____)

TCTAAGCTTGCGGCCGCGACCGGCCAGGTGCAIGCGCCGCTCGTCATCCC
50-MER

15

Example 2

Preparation of Vectors

Vectors had the general structure shown in Figure 6.

A. Plant transformation vector series

20

Initial expression vectors were based on **pCOR105** [rice actin promoter - McElroy et al. MGG 231:150-160 (1991)] (Figure 7). pCOR105 Not and SstI sites were first destroyed [cut with NotI and SstI, followed by heat inactivation and T4 DNA polymerase treatment in the presence of dNTPs] using standard
25 methods as described in Maniatis et al. or following the manufacturer's instructions for enzymes to simplify subsequent Not cassette manipulation and allow use of unique Sst site (see below).

The *nos* terminator from **pMA406** (Ainley & Key (1990) PMB 14:949-60) was amplified by PCR using primers TER5 and TER3 to generate a fragment
30 with the following sequence (SEQ ID NO: ____):

(Pst1) (Not 1)
 (AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG
 AAGCAGATCGTTCAAACATTTGGCAATAAAGTTTCTTAAGATTGAATCCTGTT
 5 GCCGGTCTTGCGATGATTATCATATAATTTCTGTTGAATTACGTTAAGCATGT
 AATAATTAACATGTAATGCATGACGTTATTTATGAGATGGGTTTTATGATTA
 GAGTCCCGCAATTATACATTTAATACGCGATAGAAAACAAAATATAGCGCGC
 AAAGTAGGATAAATTATCGCGCGCGGTGTCATCTATGTTACTAGATCGATA
 AGCTT CTA GATCT (where K=G or T)
 10 (HindIII) (XbaI)

A redundancy in the TER5 primer (GCGKAA) creates fragments having either a stop codon (TAA) or glutamate codon (GAA) in one reading frame. The glutamate codon is in frame with a downstream KDEL motif.

15 The fragment and modified pCOR105 vector were cut with PstI and XbaI, according to manufacturers instructions, relevant fragments gel-purified, ligated with T4 DNA ligase and transformed into *E. coli*. Resulting clones were then sequenced to establish which TER5 alternatives were present.

Initial FAE expression vectors were then constructed from these vectors
 20 by inserting FAE-S5/FAE-3 PCR products (T4 DNA polymerase 'polished' in the presence of dNTPs, purified and digested with NotI, cloned into EcoRV and NotI digested vector) or FAE-N5/FAE-3 PCR products (purified and NotI digested, cloned into NotI digested and calf intestinal alkaline phosphatase treated vector).

The initial pCOR105-*nos* terminator clones were also modified by the
 25 addition of ALE-5/ALE-3 PCR products (encoding wild-type and modified barley aleurain signal peptides, see below for details). The products were 'polished' with T4 DNA polymerase in the presence of dNTPs, purified and cut with NotI, then cloned into EcoRV and NotI digested vectors. Addition of the ALE sequences creates a series of vectors which can express a reading frame inserted at the
 30 NotI or NcoI sites as a fusion to the barley aleurain signal, with or without vacuolar targeting motif, and with or without an ER retention motif. HindIII sites flanking the translation initiation codon and transcriptional terminator allow easy

movement of transcription units between expression vectors providing different promoter sequences. (See Figure 8 depicting the generic ALE-TER vector.)

Vector sequences were confirmed by sequencing. Two artifacts were found. Firstly, the redundant codon in TER5 was found to be AAA in one clone, which was subsequently used as the source of all KDEL fusions (the peptide sequence is KPLKDEL, rather than EPLKDEL as designed). See Figure 9. Secondly, an additional base is found at the site of the redundant codon in one clone, creating a frameshifted terminal peptide (ETTEG, Figure 10) which was used as a control in some constructs.

Exploitation of the modular arrangement of signal peptides in the above vector series allowed various combinations of FAE and targeting motifs to be created using standard molecular biology procedures (i.e., restriction digest, purification of relevant fragments and ligation as appropriate). For example, the NotI fragment containing the FAE reading frame was inserted into the NotI site of the frameshifted clone described above to create vector **pTP3.1**. The native *Aspergillus* COOH-terminus was inserted into a FAE-S5/FAE-3 clone as a SphI (T4 DNA polymerase polished) – NcoI fragment from the FAE genomic clone (replacing the NotI (T4 DNA polymerase polished) – NcoI fragment), creating vector **pTP4a2**, which then encodes the entire, unmodified, *Aspergillus* FAE. Replacement of the SalI/XbaI fragment of pTP3.1 with that of pTP4a2 then created **pTP11.1**, which encodes FAE with a native *Aspergillus* COOH-terminus but a barley aleurain N-terminal signal.

Briefly, other vectors made in this series were; **pTP8.5**, the FAE NotI fragment inserted into the NotI site of an ALE-frameshifted COOH-terminus construct, aleurain N-terminus; **pTP5.1**, replacement of the native *Aspergillus* COOH terminus with a KDEL peptide (NotI/XbaI fragment exchange), *Aspergillus* N-terminal signal retained; **pTU4.4**, BamHI fragment of pTP11.1 replaces BamHI fragment of pTP5.1, creates FAE reading frame fused to heterologous N- and C-termini (aleurain signal and KDEL).

Vectors in which the aleurain vacuolar targeting motif NPIR was replaced by NPGR (found to be inactive in some plant assays) were created by replacing an EcoRV/NotI fragment with ALE PCR product which had been cut with *AccI*

(T4 DNA polymerase polished) and NotI (vectors pTT5.5 and pTT5.14, Aspergillus COOH-terminus). The BamHI fragment of pTT5.5 was used to replace that of pTP5.1 to produce pTU5, creating an FAE reading frame fused to heterologous N- and C-termini (NPGR modification of aleurain signal and KDEL).

- 5 The aleurain signal was also modified by PCR mutagenesis to remove the vacuolar targeting NPIR motif in its entirety (directed by primer ALECUT, which contains a NotI site to allow exchange of BglII/NotI fragments). NPIR deletion was created in this way in pTP11.1 (creating pUA4.4), and in pTP5.1 by exchange of BamHI fragments with pUA4.4 (creating pUG4).

- 10 Finally, PCR mutagenesis, using overlap of fragments generated by primers GLY3 and GLYB, was also used to alter a potential glycosylation site (asparagine codon changed to aspartate, as carried out for example in Chen, H.M., C. Ford & P. J. Reilly (1994) Biochem J 301 275-281 Substitution of asparagine residues in Aspergillus awamori glucoamylase by site-directed
15 mutagenesis to eliminate N-glycosylation and inactivation by deamidation; see sequence data for exact change, vector pTP10.1).

PCR primers

- 20 **TER-5 (SEQ ID NO:___)**
AGACTGCAGACCATGGCGGCCGCGKAACCACTGAAGGATGAGCTGTAAAG
AAGCAGATCGTTCAAACATTTG 72-MER (The KDEL stop codon is underlined.)
TER-NOT (SEQ ID NO:___)
AAGACTGCAGACCATGGCGG 20-MER
25 **TER-3 (SEQ ID NO:___)**
AGATCTAGAAGCTTATCGATCTAGTAACATAGATGACACC
ALECUT (SEQ ID NO:___)
CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC
GLYB (SEQ ID NO:___)
30 GAGGGTGTATTCGGTATCGAGTTGCAGGTTCTGTATC
GLY3 (SEQ ID NO:___)
CTCGATACCCATTACACCCTCACGCCTTTCTGA

35 *B. Construction of different promoter vectors*

Various promoters were used to optimize expression and to establish constitutive, heat-shock inducibility and senescence enhancement.

i. Rice actin promoter and 1st intron

Initial vectors (Figures 11 and 12) were constructed from pCOR105 which was subsequently found to contain a 5bp deletion relative to the published sequence which destroys the *AccI* site (GTAGGTAGAC, deleted bases underlined) and may affect splicing at the adjacent 3' site. The original rice actin sequence in this region (GTAGGTAG) was therefore restored using
 5 oligonucleotide NCO-ACT (CTCACCATGGTAAGCTTCTACC TACAAAAAGCTCCGCA) by replacing the *BglII*/*HindIII* fragment with a PCR product, to produce vector **pPQ10.1**.

A rice repetitive element is present in the upstream region of the actin promoter used in pCOR105; as this may have unpredictable effects on vector
 10 expression it was removed from pPQ10.1 by deletion of the *KpnI*/*EcoRI* fragment (end-filled with T4 polymerase and ligated following digest, restoring *EcoRI* but not *KpnI*), to produce vector pGT6. The *HindIII* fragment containing the FAE reading frame and nos terminator of pTP3.1 (see Example 2A) was then inserted
 15 into pGT6 to produce construct **pJO6.3**.

ii. Soybean heat-shock promoter

A soybean heat shock promoter from a 23kD HSP was obtained from pMA406 (Ainley & Key (1990) PMB 14:949-60). This promoter when fused to β -glucuronidase (Jefferson et al 1987 EMBO J 6:3901-3907) had previously been
 20 shown to be inducible by a 10°C heat-shock and show stable expression for 24-48 hours (data not shown). β -glucuronidase fusions are a sensitive and versatile fusion marker in higher plants. The construction of the co-integration HS vectors is given below.

25

iii. Senescence enhanced expression (See1) promoter from Lolium multiflorum

The promoter and signal sequence (including NPIR motif) of the LSee1 gene was amplified from *Lolium multiflorum* cv Tribune with oligonucleotides SEE-NCO and SEE-VAC, and cloned as an *Asp718/NotI* replacement of the promoter region of
 30 vector pTP11.1. Following sequencing to screen for PCR artifacts, one of three identical clones was chosen (pUB8.11).

The See1 promoter from maize has been cloned previously and has

EMBL accession number is AX050343. See WO0070061.

The *Lolium* version of See1 was also cloned previously (Qiang Li (2000) Studies on leaf senescence and its genetic manipulation in *Lolium multiflorum* PhD Thesis University of Wales, Aberystwyth) and has been shown to be senescence inducible when used to drive both β -glucuronidase and the *Agrobacterium* ipt gene.

An apoplast-targeted derivative was constructed by amplifying the Potato Protease Inhibitor (PPI) motif with primers PPI-AP6 and SEE-ATG, and cloning the product as an NgoMIV/NotI fragment into pUB8.11 (NgoMIV partial digest), to produce vector pJQ5.2. This vector has both the senescence induced promoter and the apoplast target sequence with the gene to be expressed inserted downstream of the apoplast sequence.

PCR Primers

SEE-VAC (SEQ ID NO: __)

AACCATGGCGGCCGCGCGCTCGGTGACGGGCCGGAT

SEE-NCO (SEQ ID NO: __)

TTCGGTACCATGGCCAGGTATAATTATGG

SEE-ATG (SEQ ID NO: __)

CTGCGCCGGCGAGATGGMCGTGACACAAGGAG

C. Construction of targeting sequences

In order to examine whether or not the localization of the enzyme would have an effect on the phenolic acid content of the cell wall various signal sequences were utilized. The targeting sequences were added either to the N-terminus or to the C-terminus of the gene of interest.

1. N-terminal signal sequences

Six N-terminal signal sequences were utilized:

- (a) The native *Aspergillus* end of FAE, plus excretion signal [apoplast localisation]

This is from the original clone and has the peptide sequence:

MKQFSAKHVLAVVVTAGHALAASTQGI.

- (b) The mature *Aspergillus* end, with no excretion signal [cytoplasmic localisation]

Peptide sequence is MAAASTQGI (underlined motif is common to all constructs). Truncation of the signal sequence in (a) above was carried out by PCR with mutagenic primer FAE-N5.

- 5 (c) The barley aleurain signal, including intact NPIR motif [vacuole localisation]

The barley aleurain vacuolar signal sequence (See Figure 13; Swissprot database accession number P05167) was derived entirely from overlapping primers (ALE-5, ALE-3, ALE-CUT ALE-CAP-5 and ALE CAP-3). Following primer annealing at 37°C and extension with T4 DNA polymerase in the presence
 10 of dNTPs according to manufacturers instructions, PCR with flanking primers ALE-5 and ALE-3 was carried out. The product was 'polished' with T4 DNA polymerase, purified, digested with NotI and cloned into EcoRV/NotI digested pCOR105-nos terminator vector (see above). ALE-3 contains redundancies so that clones encoding NPIR or NPGR motifs may be recovered. Two versions of
 15 the signal, with and without the vacuole targeting motif, were produced, to give putative vacuolar NPIR and apoplast (NPGR) signal sequences.

PCR Primers

ALE-5 (SEQ ID NO: ____)

20 GGAATTCGTAGACAAGCTTACMATGGCCACGCCCGCGTCCT 41-MER

ALE-3 (SEQ ID NO: ____)

TATCCATGGCGGCCGCGCGGTGCGGTGACGGGCCGGMTCGGGTTGGAGTC
 GGCGAA 55-MER

ALE-CUT (SEQ ID NO: ____)

25 CTAGGCGGCCGCGCGGGAGGAGGCGACGGCGAC 33-mer

ALECAP-5 (SEQ ID NO: ____)

GCGACGGCGACGGCGGCCGTGGCCAGCACGGCGAGCGCCAGGAGGAGG
 ACGCGG 54-MER

ALECAP-3 (SEQ ID NO: ____)

30 TCGCCGTCGCCTCCTCCTCCTTCGCCGACT 33-MER

(d) The barley aleurain signal, mutated to a NPGR motif [cytoplasmic localisation]

(e) The rat sialyl transferase golgi targeting motif [golgi localisation]

5 A Golgi targeting vector, pJQ3.2, was made by inserting a reading frame encoding the relevant rat sialyl transferase (RST) motif (See Figure 14. RST motif shown to function in plants by Boevink P, Oparka K, Cruz SS, Martin B, Betteridge A, Hawes C, (1998) PLANT JOURNAL 15 441-447 Stacks on tracks: the plant Golgi apparatus traffics on an actin/ER network) into vector pPQ10.1, and replacing the
10 EcoRI/NotI promoter/signal fragment of pJO6.3 with the fragment from this vector. Briefly, the RST motif was constructed by annealing oligonucleotides RST-F1A, RST-F1B, RST-F2A and RST-F2B, and amplifying the product with RST-5AD and RST-3A. This product was cloned and sequenced. Clones were found to have a
15 deletion which was corrected by PCR with RST-RPT, followed by overlap-PCR and cloning of products.

PCR primers

RST-5AD (SEQ ID NO: ____)

ACTAAGCTTAAGGAGATATAACAATGATCCACACCAACCTCAA

20 **RST-F1A (SEQ ID NO: ____)**

TTCCATGATCCACACCAACCTCAAAAAGAAGTTCTCCCTCTTCAT

RST-F1B (SEQ ID NO: ____)

AGAGTGATCACGGCGAAGAGGAGGAAGACGAGGATGAAGAGGGAGAACTTCT
TTT

25 **RST-F2A (SEQ ID NO: ____)**

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCCAA
GCCAAGGA

RST-F2B (SEQ ID NO: ____)

CATTTGGAACCTCCTTGGCTTGGAGGGTG

30 **RST-3A (SEQ ID NO: ____)**

AACCATGGCGGCCGCCATTTGGAACCTCCTTGGCT

RST-RPT (SEQ ID NO: ____)

TATAGATCTGCGTGTGGAAGAAGGGCTCCGACTACGAGGCCCTCACCTCC
AAGCCAAGGA

(f) otif [cytoplasmic localisation]

(g) The potato protease inhibitor II (PPI) apoplast motif [apoplast
5 localisation]

An apoplast targeting reading frame was designed to encode the relevant
potato protease inhibitor II (PPI) motif (See Figure 15) and cloned into **pJO6.3**, to
produce vector **pJQ4.9**. Briefly, the PPI motif was constructed by annealing
oligonucleotides PPI-AP1, PPI-AP2, PPI-AP3, PPI-AP4, PPI-AP5 and PPI-AP6, and
10 cloning this product as a HindIII/NotI fragment into vector pPQ10.1; the EcoRI/NotI
promoter/signal fragment of pJO6.3 was then replaced with the equivalent fragment
from the modified pPQ10.1 vector.

PCR primers

15 PPI-AP1 (SEQ ID NO: __)
GGAATTCGTAGACAAGCTTACMATGGMCGTGCACAAGGAGGT
PPI-AP2 (SEQ ID NO: __)
GATCAGGAGGTAGGCWACGAAGTTWACCTCCTTGTGC
PPI-AP3 (SEQ ID NO: __)
20 CCTACCTCCTGATCGTSCCTCGGCCTCCTCTTGCTCGT
PPI-AP4 (SEQ ID NO: __)
CCTTGCGCTCCACGTGCTCCATGGCGGAWACGAGCAAGAGGAG
PPI-AP5 (SEQ ID NO: __)
GTGGACGCCAAGGCCTGCACCKCGAGTGCGGCAACCTC
25 PPI-AP6 (SEQ ID NO: __)
GGAATTCGCGGCCGCCGGCAGATGCCGAAGCCGAGGTTGCCGCACT

ii. C-terminal end signal sequences

Four C-terminal signal sequences were utilized:

30 (a) Native Aspergillus end, [CTW] (vacuole and apoplast vectors)

This was derived directly from the genomic clone (see Example 1) as a
NcoI-SphI fragment (Sph end filled with T4 polymerase) which replaces the

Nco1-Not1 region of a standard actin -FAE vector (Not1 end filled with T4 DNA polymerase).

(b) Expression vector linker alone [CTW-PVAAA] (plant optimised C-terminus for vacuole, golgi and apoplast vectors)

5 CTW is the peptide sequence of the Aspergillus FAE COOH end and is here provided by oligo FAE3. In this primer the reading frame is extended to provide the additional amino acids PVAAA which are partially encoded by the Not1 site used for cloning downstream signals see c) and d) below. Some COOH amino acids /motifs may affect compartment targeting, the PVAAA
10 sequences are expected to be neutral in this respect while the native Aspergillus end may not be.

(c) Linker plus KPLKDEL [first K is primer artifact, intended to be E] {ER retention vectors)

These sequences are provided by primer TER5 introduced during PCR to
15 generate the nos terminator fragment, and identified by sequencing within a specific clone. KDEL targeting has been demonstrated in plants by Denecke et al. ((1992) EMBO J 11: 2345-2355 Plant and mammalian sorting signals for protein retention in the endoplasmic reticulum contain a conserved epitope).

(d) Linker plus ETTEG [frameshift of (c)] (loss of ER retention - vacuole
20 vectors)

These sequences are provided by primer TER5 introduced during PCR to generate the nos terminator fragment, and identified by sequencing within a specific clone (see Example 2A).

The KDEL signal is for ER retention, while others provide controls. A
25 frameshift in the TER5 region [additional A] was used in subsequent constructs to destroy the ER KDEL retention signal.

The linker used in the above C-terminal targeting sequences was PVAAA.

D. Co-integration and co-transformation vectors.

30

Co-transformation vectors

A Hygromycin resistance gene driven by a CaMV345S promoter (**pRob5**) (35S-HYG-CMV in pUC18 (modified HYG, derived from pGL2) Bilanz et al (1991)

Gene 100:247-50) was used for co-transformation experiments with **pTT3** and **pTP3.1**, **pJQ4.9**, **pJQ3.2**, **pJO6.3**, **pJQ5.2**, **pUB8.1** 1 vectors.

5 *Co-integration vectors*

1. **Actin promoter constructs** - **pTR2.22**, **pTR6.1**, **pTR8.1**, **pTR9.4**, **pTR7.1**, **pTT5.5** and **5.1**.

10 The CAMV35S-hyg region from **pAJEB64TCA** [a plant expression vector constructed by Andy Bettany at IGER containing CaMV-HYG from **pTRA151** (Zheng et al 1991 Plant Physiol 97:832-835) (CaMV35S-HYG-tnl terminator as clonable cassette in **pUC4**) cloned into KpnI site of **pCOR105**] was added as a HindIII fragment at the KpnI site (T4 polymerase blunt) of **pTP4a2**, in divergent orientation
15 to FAE to create **pTR2.22**. The FAE/Nos HindIII fragment of this vector was replaced as follows in co-expression vectors. From **pTP5.1** for **pTR6.1**, from **pTP10.1** to **pTR8.1**, from **pTP11.1** to **pTR9.4**. Signal sequences of FAE in **pTR2.22** were replaced as HindIII/BglII fragments in **pTR7.1** (fragment from **pT09.1**). PCR products (**ALE5/ALE-G**) was digested with **Acc1** and T4 polymerase, polished,
20 followed by **Not1** digest and cloning into **EcoRV/ Not1** digested **pTR2.22** to give clones **pTT5.5** and **5.1**.

PCR primer

ALE-G

25 TATCCATGGCGGCCGCGCGGTCCGGTGACGGGCGGCCCGGGTTGGAGTC
GGCGAA

2. **Actin promoter constructs** -**pUF1**, **pUA1K3**, **pUH4**, **pUH5**, **pUH6**, **pUH7**, **pUH8**, **pUH9**.

30 The HygR gene from **pAJEB64TCA**, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of **pTP3.1**, to give **pHOX3**. For ease of cloning the downstream HindIII site was destroyed to create **pUA1K3** and replacement of the FAE/Nos terminator HindIII fragment in this vector
35 was carried out as follows. From **pTP5.1** for **pUF1**, from **pTP11.1** for **pUH4**, from **pTP8.5** for **UH5**, from **pTT5** for **pUH6**, from **pUA4.4** for **pUH7**, from **pTU5** for **pUH8** and from **pUG4** for **pUH9**.

3. Heat-shock promoter constructs - pUH10, pUH12, pUC5.11.

5 A co-transformation vector in which FAE is expressed from the soybean heat shock promoter was made by first modifying pMA406 to remove the nos terminator (BglII linearised and gel purified, KpnI digested, T4 DNA polymerase polished in the presence of dNTPs and recircularised), and then inserting the FAE HindIII fragment from pTP11.1, creating pTT3.1, which encodes the full aleurain signal and the native
10 *Aspergillus* COOH-terminus.

Following assays of various constructs, co-integration vectors were constructed with FAE and HygR genes arranged in tandem.

The HygR gene from pAJEB-64-TCA, driven by the CaMV promoter, was first cloned as an end-filled HindIII fragment at the end-filled XbaI site of pTP3.1, to give pHOX3 and subsequently excised as a HindIII/SacI fragment (partial SacI
15 digest, relevant sites found in flanking pTP3.1 sequences) which was cloned into the HindIII/SacI sites of pMA406, in tandem orientation (vector pUH1a20). FAE sequences were then cloned into the HindIII site of pUH1a20 downstream of the heat-shock promoter (HindIII fragment from pTU5 for pUH10, HindIII fragment from pTT5 for pUH12). A pTP3.1 derivative was made by cloning the
20 CaMV/HygR HindIII cassette from pAJEB-64-TCA in tandem orientation downstream of the FAE gene in pTP3.1, inactivating the middle HindIII site by partial digestion and end-filling, and excising the combined FAE/HygR cassette as a single HindIII fragment, which was inserted at the HindIII site in pMA406 to
25 produce pUC5.11.

Example 3

Transformation of Plant Cells

Eight to ten weeks old embryogenic *F. arundinacea* and *L. multiflorum*

30 suspension cultures were bombarded either with a single co-integration plasmid DNA vector containing FAE and hyg resistance genes, or with a co-transformation vector containing FAE and with plasmid pROB5 conferring hygromycin resistance (CAMV35S-hpt- nos) using a Particle Inflow Gun (PIG) (Finer et al. (1992) Development of the particle inflow gun for DNA delivery to

plant cells Plant Cell Reports 11:323-328) and 1.5-3.0 μm gold particles as in Dalton *et al* (Dalton *et al.* (1999) Co-transformed diploid *Lolium perenne* (*Perennial ryegrass*), *Lolium multiflorum* (Italian ryegrass) and *Lolium temulentum* (Darnel) plants produced by microprojectile bombardment. Plant Cell Reports. 5 18: 721-726) and Kuai *et al* (Regeneration of fertile transgenic tall fescue (*Festuca arundinacea*) plants with a stable highly expressed foreign gene. Plant Cell Tissue and Organ Culture (1999) 58:149-154). Transformants were selected with hygromycin (25 to 50mg/l) over a 10-12 week selection period at 25°C under continuous white fluorescent light ($60 \mu\text{E m}^2 \text{s}^{-1}$) and plants regenerated via 10 somatic embryogenesis as in Dalton *et al* 1999, *supra*. Regenerated plants were screened for FAE activity on transfer to soil and expressing plants grown to maturity in a containment growth room at 18°C under 16h fluorescent lights ($350 \mu\text{E m}^2 \text{s}^{-1}$). Mature plants (6-8 weeks old) were re-assayed for FAE activity and fresh tissue harvested for Southern, Northern and Western analysis, and for self 15 digestion analysis. The remaining tissue was freeze dried and powdered for cell wall structure analysis, In vitro-dry matter digestibility (IVDMD) determinations and for in-vitro gas production determinations of rates of tissue digestion.

Example 4

Targeting of Expression Product

20 To verify that the targeting sequences are effective in delivering the gene the targeting sequences were operably linked to a green fluorescent protein GFP. The vector constructs are shown in Figure 16. Cells were transformed by particle bombardment as in Example 3. Localization of the GFP could be 25 visualized under a microscope 1 day after bombardment (i.e., shooting). See Figure 16.

Example 5

FAE1 activity

30 Plants regenerated from transformed cells showed FAE activity in all plant tissues tested. Cells were transformed as above under the direction of the ER and APO targeting sequences. FAE activity in transformed *Festuca arundinacea*

leaves of different ages was elevated compared to control (untransformed) plants. See Figures 17 and 18.

Similar results were seen with *Lolium mutiflorum* leaves at different ages transformed as above under the direction of vacuolar, ER and APO targeting sequence. See Figures 19 and 20.

FAE expression under a heat shock promoter can also be induced. (Data not shown.)

Thus, we have demonstrated FAE expression in *Festuca* and *Lolium* leaves under constitutive and HS promoters with effective FAE targeting to the vac, ER and apo.

FAE assay

FAE activity was determined in soluble extracts of fresh (or frozen at -70°C) leaves or cell cultures (0.5g) with 0.1M NaAc, pH 5.0 extraction buffer. Extracts were incubated with 24mM EF (ethyl 4-hydroxy-3-methoxycinnamate) or 1% FAXX as substrate, at 28°C for 24hrs and FAE activity calculated as the amount of ferulic acid released. FAE activity was also determined by measuring the release of monomeric and dimeric ferulic acid from self-digested leaf or cell culture samples. Fresh, or frozen, leaves or cell cultures (0.5g) were ground in 0.1M NaAc, pH5.0 extraction buffer in the presence and absence of xylanase (1000U GC140/sample) without added substrate and incubated at 28°C for 72hrs. Following incubation, and centrifugation, soluble extracts were loaded onto an activated reverse phase C18 μ Nova sep-pak column (Waters), eluted with 100% MeOH and the MeOH sample analysed by HPLC.

25

Example 6

Chemical Analysis of Cell Wall Extracts

Ester bound compounds were extracted from freeze dried powdered leaves or cell cultures (50 -100mg) with NaOH (5ml of 1M) followed by incubation at 25°C for 23hrs under N₂. After centrifugation and acidification of the soluble extract with concentrated HCl, the extracted phenolics were loaded onto an activated reverse phase C18 μ Nova sep-pak column (Waters) and eluted with 100% MeOH. and the MeOH sample analysed by HPLC.

HPLC was carried out with methanol: 5% acetic acid either with a 35-65% MeOH gradient in 15min (FAE assay) or with a 30-70% MeOH gradient in 25 min (monomer and dimer cell wall components) at 2ml/min on a µNova Pak C18 8x10 RCM (Waters). Extracts were detected and quantified with a diode array detector
5 (240-400nm Waters 996PDA) monitored at 280nm for aldehydes and 340nm for hydroxycinnamic acids. .

Levels of esterified monomeric and dimeric hydroxycinnamic acids in *Festuca arundinacea* plants expressing FAE under VAC, and ER and APO targeting sequences are reduced compared to control (untransformed) plants.
10 The results can be seen in Figure 21 and 22, respectively. Thus, we show where this does not result in reduced cell wall phenolics in growing plants with vac targeting but does result in lower phenolics with ER and apo targeting. In addition,

Levels of esterified monomeric and dimeric hydroxycinnamic acids in
15 *Festuca arundinacea* plants expressing FAE are not significantly reduced when FAE is VAC, targeting (Fig 21) which is as predicted for correct vacuolar targeting, but are significantly reduced, as predicted, in some plants when FAE was ER and APO targeted, compared to control (untransformed) plants. The results can be seen in Figure 22.

20

Example 7

In vitro dry matter digestibility. (IVDMD)

The *in vitro* dry matter digestibility (IVDMD) was estimated on 1.0 g dry weight of leaf or cell culture tissue using the pepsin/cellulase method of Jones
25 and Hayward (The effect of pepsin treatment of herbage on the prediction of dry matter digestibility from solubility in fungal cellulase solutions. Journal of the Science of Food and Agriculture (1975) 26:711-718).

We show that the presence of FAE in the plants results in higher digestibility of the leaves. This may be due to internal FAE activity acting on
30 normal cell walls with vacuole located FAE and to both FAE activity and the lower cell wall crosslinking with ER and apo targeted FAE (as also found with cell cultures).

End point digestibility as determined by IVDMD were higher in leaf tissue of some transformed plants of *Festuca* expressing FAE, compared to control (untransformed) plants. Examples are shown where vacuolar, ER or apoplast targeted FAE under a constitutive actin promoter have been effective at increasing IVDMD. Similar results were obtained with in leaves of *Lolium*, but were less pronounced.

The results can be seen in Figures 23 and 24.

Example 8

In vitro gas production measurements

In each experiment, 1.0-g samples of freeze dried powdered leaf tissue or cell culture were fermented in three 165-ml capacity serum bottles according to the pressure transducer technique of Theodorou et al. (Theodorou et al. (1994) A new gas production method using a pressure transducer to determine the fermentation kinetics of ruminant feeds. *Animal Feed Science and Technology* 48: 185-197). Grab samples of rumen-digesta were taken at 8.00 h before the morning feeding from fistulated wethers fed grass hay, and transported to the laboratory in a pre-warmed (39°C) vacuum flask. The microbial inoculum and culture media were prepared as described by Theodorou et al. (1994). Each serum bottle received 10 ml of microbial inoculum, 85 ml of buffer and 4 ml of reducing agent.

At the end of the incubation period, (144h) the contents of each serum bottle were filtered through pre-weighed sintered glass funnels and freeze dried to constant weight. Dry matter loss was calculated as the difference between the dry weight of the sample pre- and post-incubation. Additionally, the concentration of volatile fatty acids (VFA) in the liquid fraction of the culture media at the end of the 144-h incubation period was determined by gas chromatography. A Chrompack CP 9000 chromatograph fitted with an automatic sampler (Chrompack 911) and a flame-ionisation detector, linked to a Dell PC with A1-450 integration software, was used for VFA quantification.

Gas production data were fitted to the model of France et al. (France, J., Dhanoa, M.S., Theodorou, M.K, Lister, S.J., Davies. D.R. and Isac, D. 1993. A model to interpret gas accumulation profiles associated with *in vitro* degradation of ruminant feeds. *Journal of Theoretical Biology*. 163: 99-111.) using the MLP (Ross,

G.J.S. 1987: *MLP, Maximum Likelihood Program Version 3.08*. Oxford Numerical Algorithms Group) package. The equation is in the form, $Y = A\{1 - e^{[-b(t-T) - c\sqrt{t-T}]}\}$, where Y is the cumulative gas production (ml), A is the asymptote (i.e. gas pool), T is lag time, and b (h^{-1}) and c ($h^{-0.5}$) are decay rate constants. A combined fractional rate (h^{-1}) of gas production (μ) was calculated as, $\mu = b + c/2\sqrt{t}$, where t is the incubation time (h).

It can be seen for *Festuca arundinacea* (denoted as BN in Figure 25) that cell cultures have a higher rate of digestion and cumulative gas production in the presence of FAE and that the addition of an exogenous xylanase further enhance the availability of fermentable carbohydrates. Similar results are found in FAE expressing cultures without added FAE. Fermentation rates are further increased compared with controls by the addition of exogenous FAE or xylanase as these cultures expressing FAE have a reduced cell wall phenolic composition to controls Figures 26-28.

15

Example 9

FAE & xylanase transformed plants

Addition of exogenous xylanase (GC140) greatly increased FAE mediated release of phenolics from *Festuca* and *Lolium* leaves expressing *A. niger* FAE. See Figures 29-31 which show that phenolic release from leaf cell walls is increased in all FAE expressing plants on cell death and this is stimulated by xylanase irrespective of the targeting. Therefore expression of a fungal xylanase in plant cells is tested.

The FAE expression cassette is modified to comprise a fungal xylanase gene (either *T. reesei* or *A. niger*) to yield a FAE-xylanase expression cassette. The FAE-xylanase expression cassette is used to transform plant cells in a manner similar to those described in Example 3. The transformed cells are allowed to grow and are selected on an appropriate medium. The enzymes so expressed increase the availability of fermentable carbohydrates to a greater extent than the FAE expression cassette.

30

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity and understanding, it will

be obvious that certain changes and modifications may be practiced within the scope of the appended claims.

What is claimed:

1. A transgenic plant comprising an expression cassette comprising a promoter operably linked to a ferulic acid esterase encoding polynucleotide.
5
2. The plant of claim 1, wherein the polynucleotide is derived from *Aspergillus niger*.
3. The plant of claim 2, wherein the polynucleotide is FAE ! from *Aspergillus niger*.
- 10 4. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with an altered glycosylation site.
5. The plant of claim 3, wherein the polynucleotide encodes a ferulic acid esterase with a substitution so that glycosylation is altered.
- 15 6. The plant of claim 3, wherein the polynucleotide further comprises a polynucleotide that encodes CTWPVAAA at the 3' end.
7. The plant of claim 3 wherein sub-optimal codons are modified to *Triticum spp.* preferred codons.
8. The plant of claim 1, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by sexual reproduction.
- 20 9. The plant of claim 1, wherein the promoter is an inducible promoter.
10. The plant of claim 9, wherein the promoter is a senescence promoter.
11. The plant of claim 9, wherein the promoter is a heat shock
25 promoter.
12. The plant of claim 1, wherein the promoter is a constitutive promoter
13. The plant of claim 1, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the

polynucleotide.

14. The plant of claim 13, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

15. The plant of claim 14, wherein the polynucleotide is derived
5 from the signal sequence of a vacuolar targeted gene

16. The plant of claim 15, wherein the targeted gene is a barley aleurain gene.

17. The plant of claim 15, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum or apoplast
10 signal sequence.

18. The plant of claim 15, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted senescence gene.

19. The plant of claim 18, wherein the senescence gene is a *Lolium Sec1* signal sequence.

20. The plant of claim 13, wherein the polynucleotide is derived
15 from the signal sequence of a golgi targeted gene.

21. The plant of claim 20, wherein the targeted gene is a rat sialyl transferease signal sequence.

22. The plant of claim 13, wherein the polynucleotide is derived
20 from the signal sequence of an apoplast signal sequence.

23. The plant of claim 22, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

24. The plant of claim 16, wherein the polynucleotide is derived from *Solanum tuberosum*.

25. The plant of claim 13, wherein the polynucleotide sequence
25 is downstream of the C-terminus of the ferulic acid esterase polynucleotide

26. The plant of claim 25, wherein the polynucleotide sequence is a KDEL sequence.

27. The plant of claim 25, wherein the polynucleotide sequence is a stop codon.

28. The plant of claim 25, wherein the polynucleotide sequence is an extension of the ferulic acid esterase reading frame to provide a linker to
5 KDEL.

29. The plant of claim 1, further comprising introduction into the plant a second expression cassette comprising a promoter operably linked to a xylanase encoding polynucleotide.

30. The plant of claim 29, wherein the xylanase encoding
10 polynucleotide is from *Trichoderma reesei*.

31. The plant of claim 29, wherein the first and second expression cassettes are present on separate plasmids.

32. The transgenic plant of claim 1, selected from the group consisting of Festuca, Lolium, Zea and Avena.

33. The transgenic plant of claim 32, wherein the plant is a
15 Festuca plant.

34. A method of controlling the level of phenolic acids in plant cell walls of a transgenic plant, the method comprising introducing into the plant an expression cassette comprising a promoter operably linked to a ferulic acid
20 esterase encoding polynucleotide.

35. The method of claim 34, wherein the polynucleotide is derived from *Aspergillus niger*.

36. The method of claim 35, wherein the polynucleotide is a FAE
1 gene from *Aspergillus niger*.

37. The method of claim 36, wherein the polynucleotide encodes
25 the ferulic acid esterase with an altered glycosylation site.

38. The method of claim 36, wherein the polynucleotide encodes the ferulic acid esterase with a substitution such that glycosylation is altered.

39. The method of claim 36, wherein the polynucleotide

comprises CTWPVAAA at the 3' end.

40. The method of claim 36 wherein sub-optimal codons are modified to *Triticum* spp. preferred codons.

41. The method of claim 36, wherein the polynucleotide
5 comprises SEQ ID NO:1.

42. The method of claim 34, wherein the introduction of the ferulic acid esterase polynucleotide into the plant is by transformation of cell cultures.

43. The method of claim 42, wherein the cell cultures are
10 regenerated to plants.

44. The method of claim 34 wherein the ferulic acid esterase polynucleotide is introduced into the plant by sexual reproduction.

45. The method of claim 34, wherein the transgenic plant is a member of a genus selected from the group consisting of *Festuca*, *Lolium*,
15 *Avena* and *Zea*.

46. The method of claim 45, wherein the transgenic plant is a member of the genus *Festuca*.

47. The method of claim 46, wherein the transgenic plant is a *Festuca arundinacea*.

48. The method of claim 34, wherein the promoter is an
20 inducible promoter.

49. The method of claim 48, wherein the promoter is a senescence promoter.

50. The method of claim 48, wherein the promoter is a heat
25 shock protein promoter.

51. The method of claim 34, wherein the promoter is a constitutive promoter.

52. The method of claim 51, wherein the promoter is an actin

promoter.

53. The method of claim 34, wherein the expression cassette further comprises a polynucleotide sequence that targets expression of the polynucleotide.

5 54. The method of claim 53, wherein the polynucleotide sequence is upstream of the N-terminus of the ferulic acid esterase polynucleotide.

55. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a vacuolar targeted gene.

10 56. The method of claim 55, wherein the targeted gene is a barley aleurain gene.

57. The method of claim 55, wherein the polynucleotide is derived from the signal sequence of a *Lolium See1* signal sequence.

15 58. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce a endoplasmic reticulum signal sequence.

59. The method of claim 55, wherein the vacuolar signal sequence of the polynucleotide is modified to produce an apoplast signal sequence.

20 60. The method of claim 54, wherein the polynucleotide is derived from the signal sequence of a golgi targeted gene.

61. The method of claim 60, wherein the targeted gene is a rat sialyl transferease signal sequence.

25 62. The method of claim 59, wherein the polynucleotide is derived from the signal sequence of a fungal apoplast signal sequence.

63. The method of claim 62, wherein the signal sequence is from *Aspergillus niger* ferulic acid esterase.

64. The method of claim 59, wherein the apoplast signal

sequence is from a potato.

65. The method of claim 53, wherein the polynucleotide sequence is downstream of the C-terminus of the ferulic acid esterase polynucleotide

5 66. The method of claim 65, wherein the polynucleotide sequence is a KDEL sequence.

67. The method of claim 65, wherein the polynucleotide sequence is a stop codon.

68. The method of claim 65, wherein the polynucleotide
10 sequence is an extension of the ferulic acid esterase reading frame to provide a linker to KDEL.

69. The method of claim 34, further comprising simultaneous introduction into the plant a second expression cassette comprising a promoter operably linked to a polynucleotide encoding a xylanase gene.

15 70. The method of claim 69, wherein the second polynucleotide is a fungal xylanase.

71. The method of claim 70, wherein the fungal xylanase is from *Trichoderma reesei*.

72. The method of claim 35, wherein the first and second
20 expression cassettes are present on separate plasmids.

73. The method of claim 1, wherein the first and second expression cassettes are present on separate plasmids.

74. A transgenic plant produced by the method of claim 34.

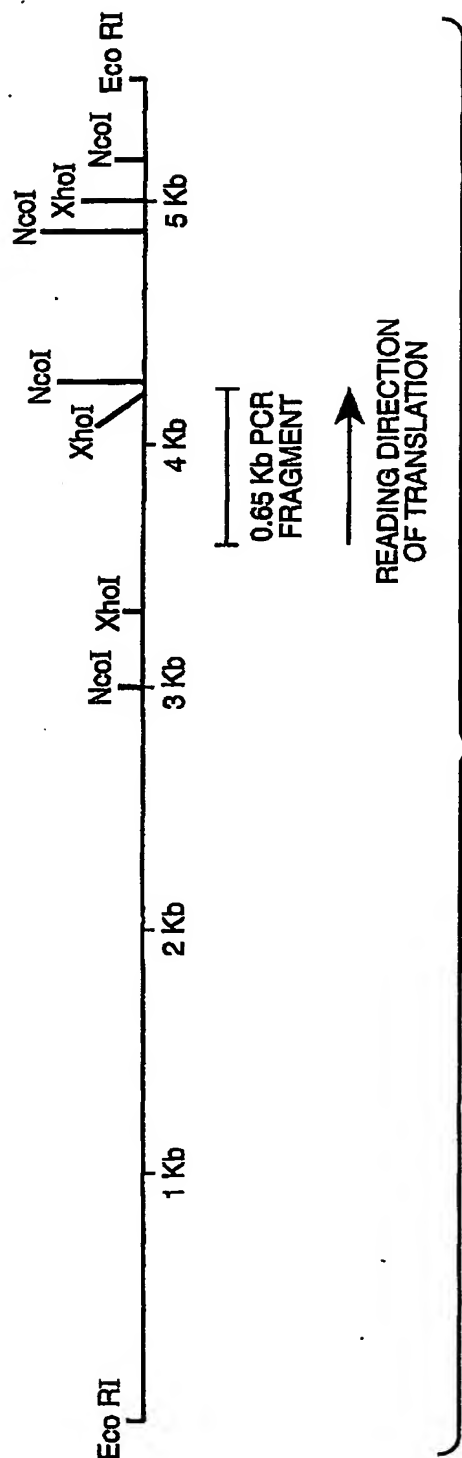
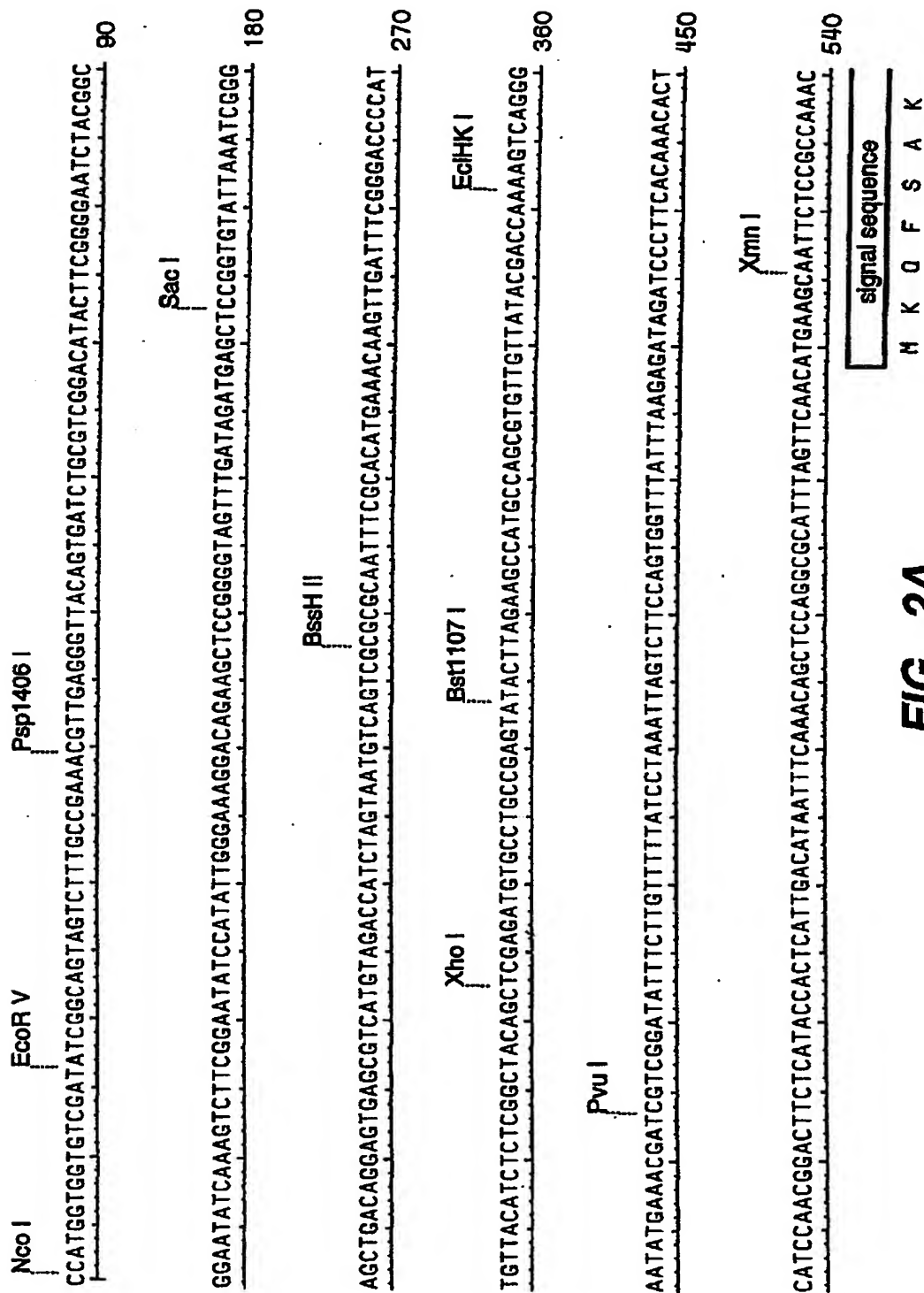


FIG. 1



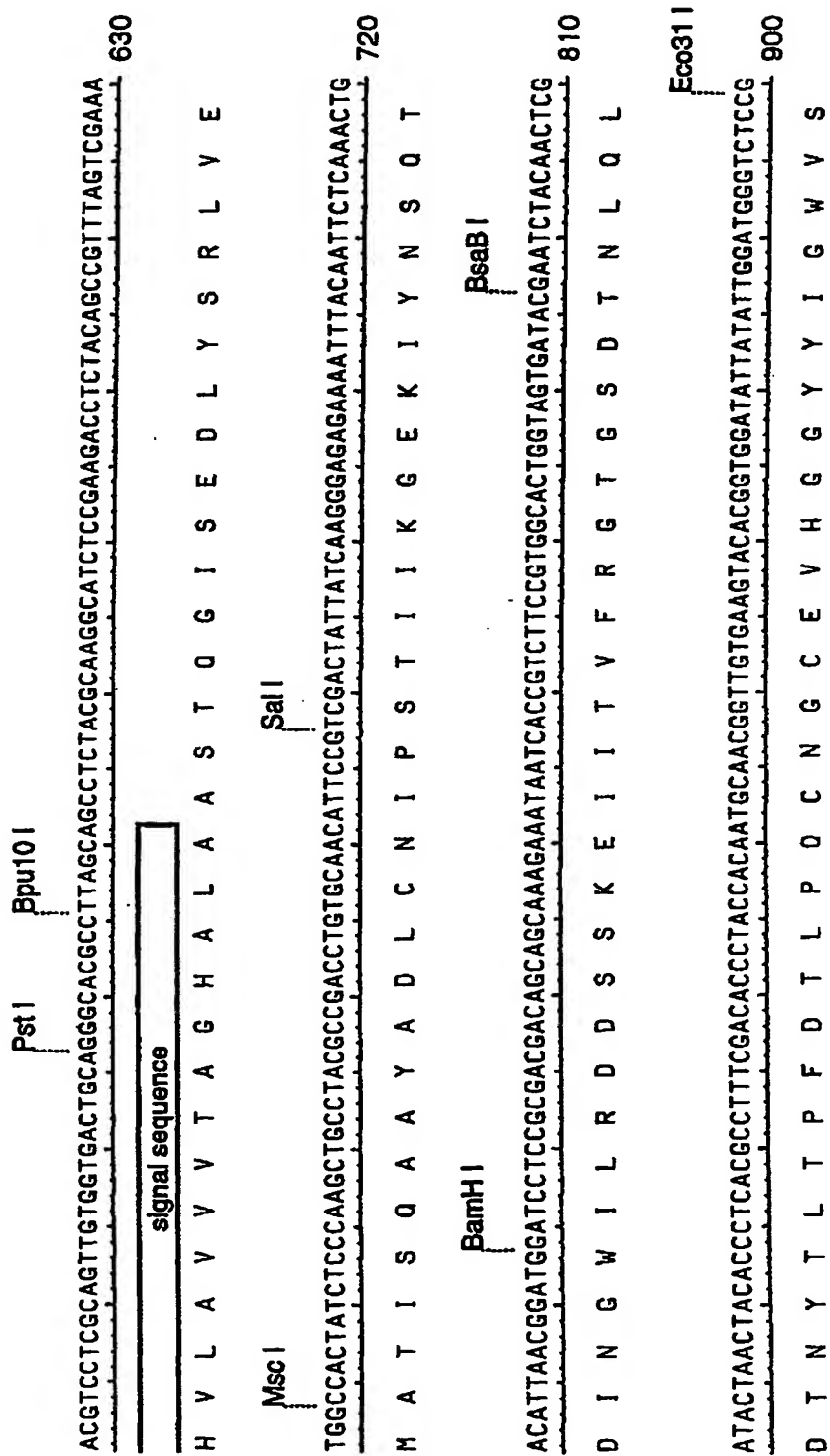


FIG. 2B

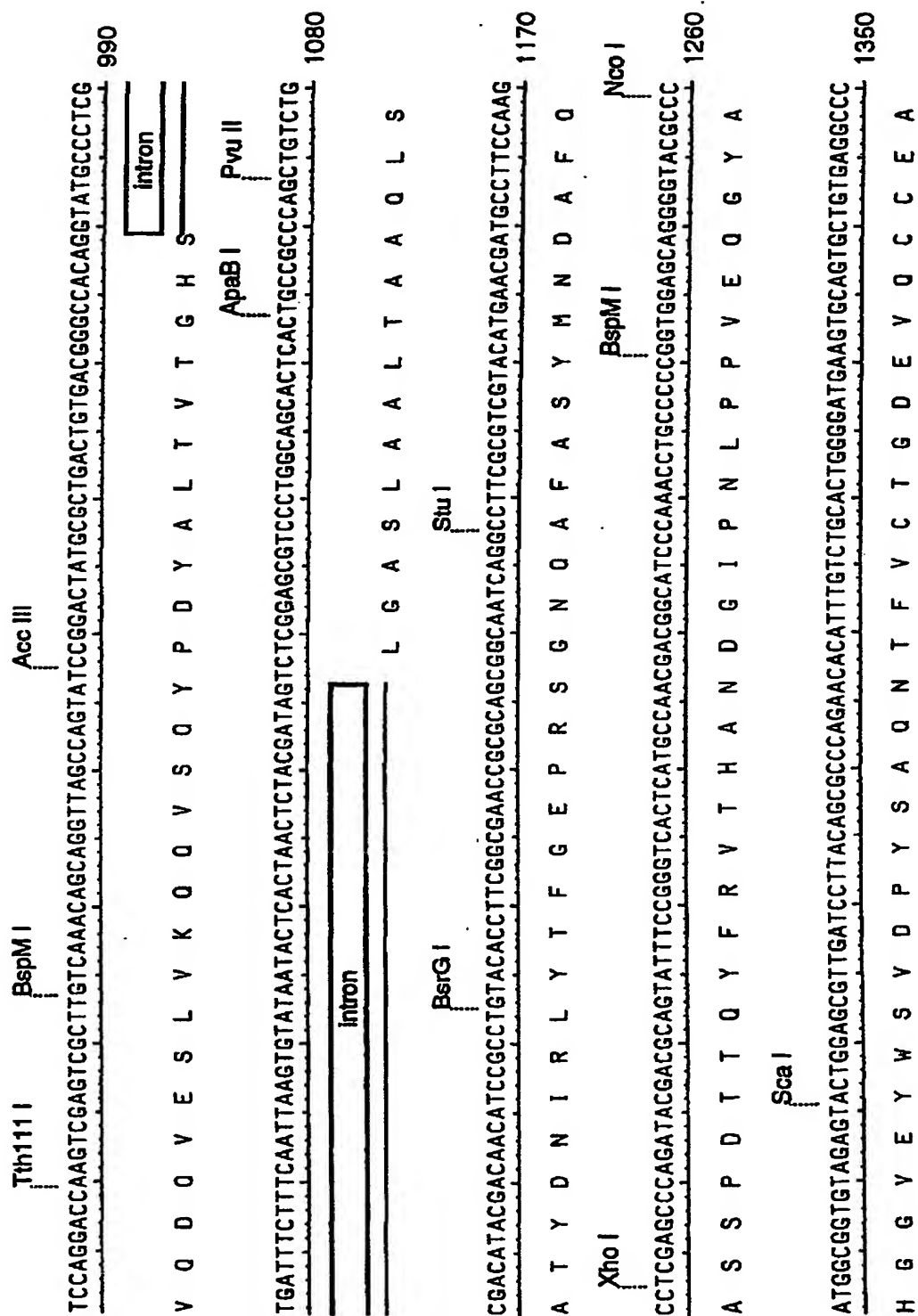


FIG.-2C

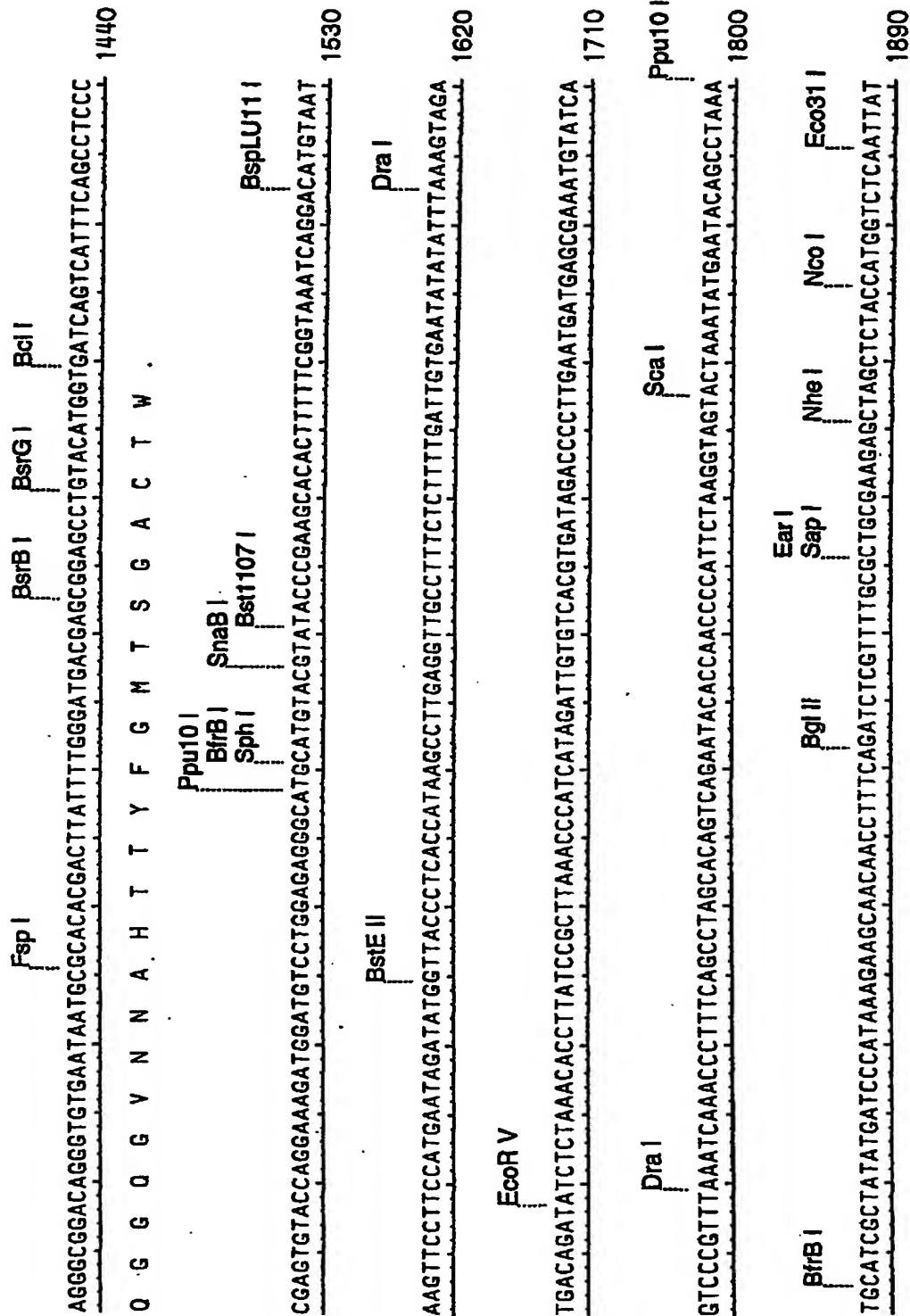


FIG._2D

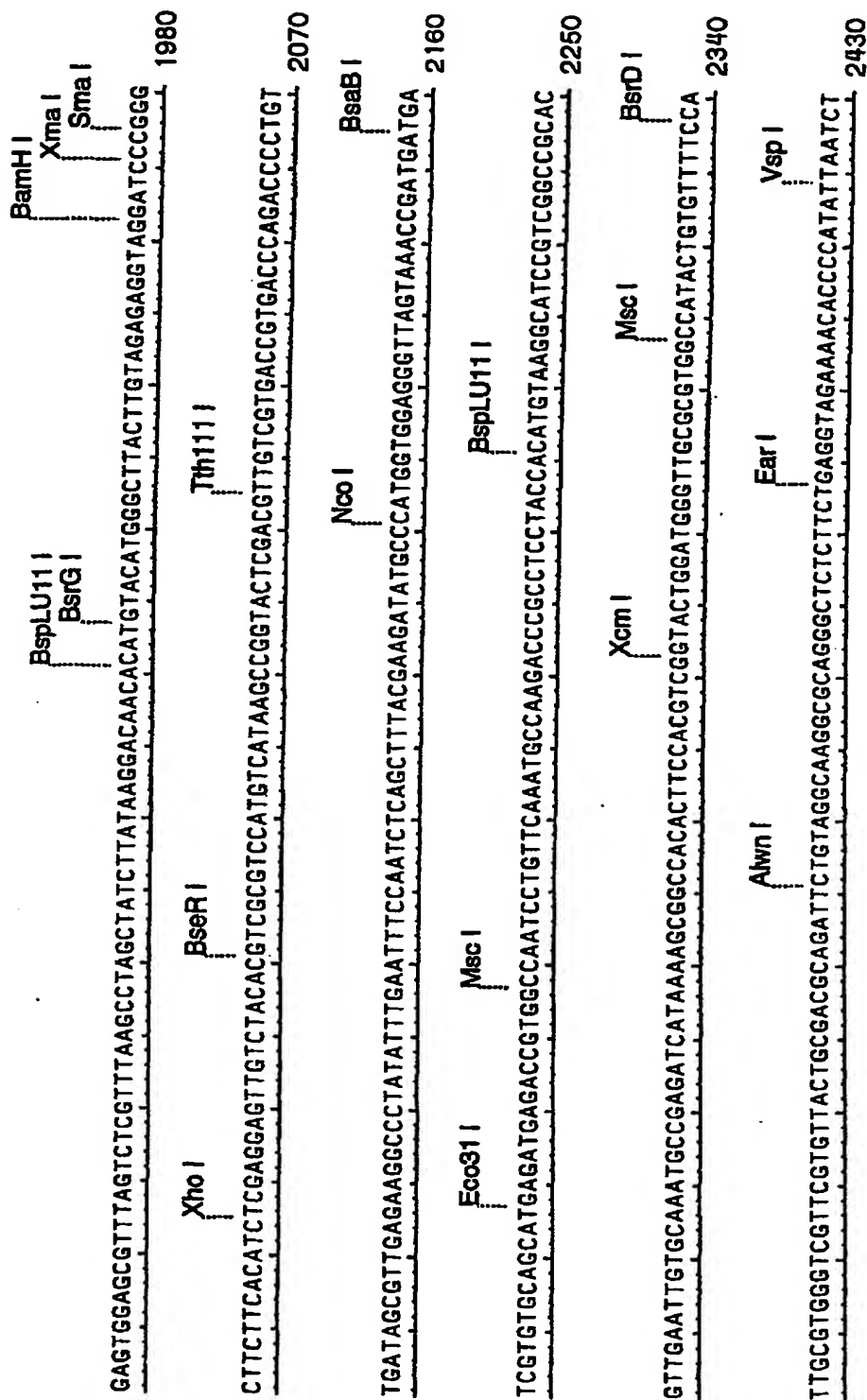


FIG. 2E

CCATGGTGGTGCATATCGGCAAGTAGTCTTTGGCCGAAACGTTGAGGGTTACAGTGATCTGCGTCGGACATACCTT
CGGGGAATCTACGGCGGAATATCAAAGTCTTCGGAATATCCATATTTGGGAAAGGACAGAAAGCTCCGGGGTAGTTT
GATAGATGAGCTCCGGTGATTTAAATCGGGAGCTGACAGGAGTGAGCGTCATGTAGACCATCTAGTAATGTCAGT
CGCGCGCAATTTCCGACATGAAACAAGTTGATTTCCGGGACCCCATTTGTACATCTCTCGGCTACAGCTCGAGATG
TGCCCTGCCGAGTATACTAGAAAGCCATGCCAGCGGTGTTGTTATACGACCAAAAGTCAGGGAATATGAAACGATCG
TCGGATATTTCTGTTTATATCCTAAATTAGTCTTCCAGTGGTTTATTTAAGAGATAGATCCCTTCACAAACACT
CATCCAACGGACTTCTCATACCACTCATTTGACATAAATTTCAAACAGCTCCAGGCGCATTTAGTTCAACATGAAGC
AATTCCTCCGCCAAACACGTCCTCGCAGTTGTGGTGAATGCTGCAAGGACCTTAGCAGCCCTCTACGCAAGGCATCT
CCGAAGACCTCTACAGCCGTTTAGTCGAAATGGCCACTATCTCCCAAGCTGCCCTACGCCGACCTGTGCAACATTC
CGTCGACTATTTATCAAGGAGAGAAAATTTACAATTTCTCAAACCTGACATTAACGGATGGATCCTCCGCGACGACA
GCAGCAAGAAAATAATCACCGTCTTCCGTGGCACTGGTAGTGATACGAACTCAAACTCGATACATACTACACCC
TCACGCCCTTTCGACACCCCTACCAAAATGCAACGGTTGTGAAGTACACGGTGGATATATATGGATGGGTCTCCG
TCCAGGACCAAGTCGAGTCGCTGTGTCAAAACAGCAGGTTAGCCAGTATCCGGACTATGCGCTGACTGTGACGGGCC
ACAGGTATGCCCTCGTGATTTCTTTCAAATTAAGTGATATAATACTCACTAATCTACGATAGTCTCGGAGCGTCCC
TGGCAGCACTCACTGCCGCCAGCTGTCTGCGACATACGAAACATCCGCCCTGTACACCTTCGGCGAACCCGCGCA
GCGGCAATCAGGCCCTTCGCTCGTACATGAACGATGCTTCCAAAGCTCGAGCCCAAGTACGACGCAATATTTC
GGTCACCTCATGCCAACGACGGCATCCCAAACCTGCCCGTGAGCAGGGGTACGCCCATGGCGGTGTAGAGT
ACTGGAGCGTTGATCCTTACAGCGGCCCAGAACACATTTGTCTGCACTGGGGATGAAAGTGCAGTCTGTGAGGCC
AGGGCGGACAGGGGTGAAATAATGCGCACACGACTTATTTTGGGATGACGAGCGGAGCCCTGTACATGGTGATCAG
TCATTTTCAGCCCTCCCGAGTGTAACAGGAAAGATGGATGTCCTGGAGAGGCGCATGCAATGTACGTATACCCGAAAGC
ACACTTTTTCGGTAAATCAGGACATGTAAATAAGTTCTTCCATGAATAGATATGGTTACCTCAACCATAGCCCTT
GAGGTTGCCCTTCTCTTTGATTTGTGAATATATATTTAAAGTAGACAGATATCTCTAAACACCTTATCCGCT
TAAACCCATCATAGATTGTGTCACGTGATAGACCCCTTGAATGATGAGCGAAATGTATCAGTCCCGTTTAAATCA
AACCCTTTCAGCTAGCACAGTCAGAAATACCAACCCCATTTCTAAGGTAGTACTAATAATAGATACAGCCCTAAA
TGCAATCGCTATATGATCCCATAAAGAAAGCAACAACCTTTCAGATCTCGTTTTCGCTGCGAAGAGCTAGCTCTAC
CATGGTCTCAATTATGAGTGGAGCGTTTAGTCTCGTTTAAAGCCTAGCTATCTTATAGGACAAACATGTACATG
GGCTTACTTGTAGAGAGGTAGGATCCCGGCTTCTTCACATCTCGAGGAGTTGTCTACACGTCGCTCCATGTCA
TAAAGCCGGTACTCGACGTTGTCGTGACCGTGACCCAGACCCCTGTTGATAGCGTTGAGAAAGGCCCTATATTGAA
TTTCCAAATCTCAGCTTACGAAAGATATGCCCATGGTGAGGGTTAGTAAACCGATGATGATCGTGTGCAGCATGA
GATGAGACCGTGCCCAATCCTGTGTTCAAATGCCAAGACCCGCTCTACCATGTAAAGGCATCCGTCGGCCGAC
GTGAAATGTGCAAAATGCCGAGATCATAAAGCGGCCACACTTCCACGTCCGTACTGGATGGGTGCGGTGGCC
ATACTGTGTTTTCATGCGTGGGTGCTTCGTGTTACTGCGACGCAATCTCTGTAGGCAAGGCGCAGGGCTCTCT
TCTGAGGTAGAAAACACCCCATATTAATCTGAATTC

FIG._3

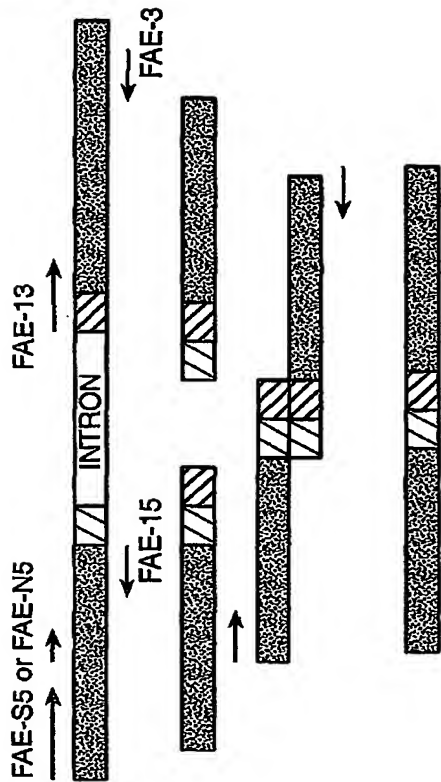


FIG._4

FAE-13 CCGGCCACGCCCTCGGGCCCTCCCTGGCGGCATC 35-mer
 FAE-15 GCGCCGAGGAGTGGCCGGTCACGGTCAGCGCGTAGTCC 40-mer

intron position in original

Y A L T V T G H S L G A S L A A L
 G G A C T A C G C G T G A C C G T G A C C G G C C A C T C C C T C G G C G C C
 C C G G C C A C C C C T C G G C G C C T C C C T G G C G G C A C T C
 Y A L T V T G H A L G A S L A A L

complement, FAE-15
 FAE-13

FIG._5

Vector Construction

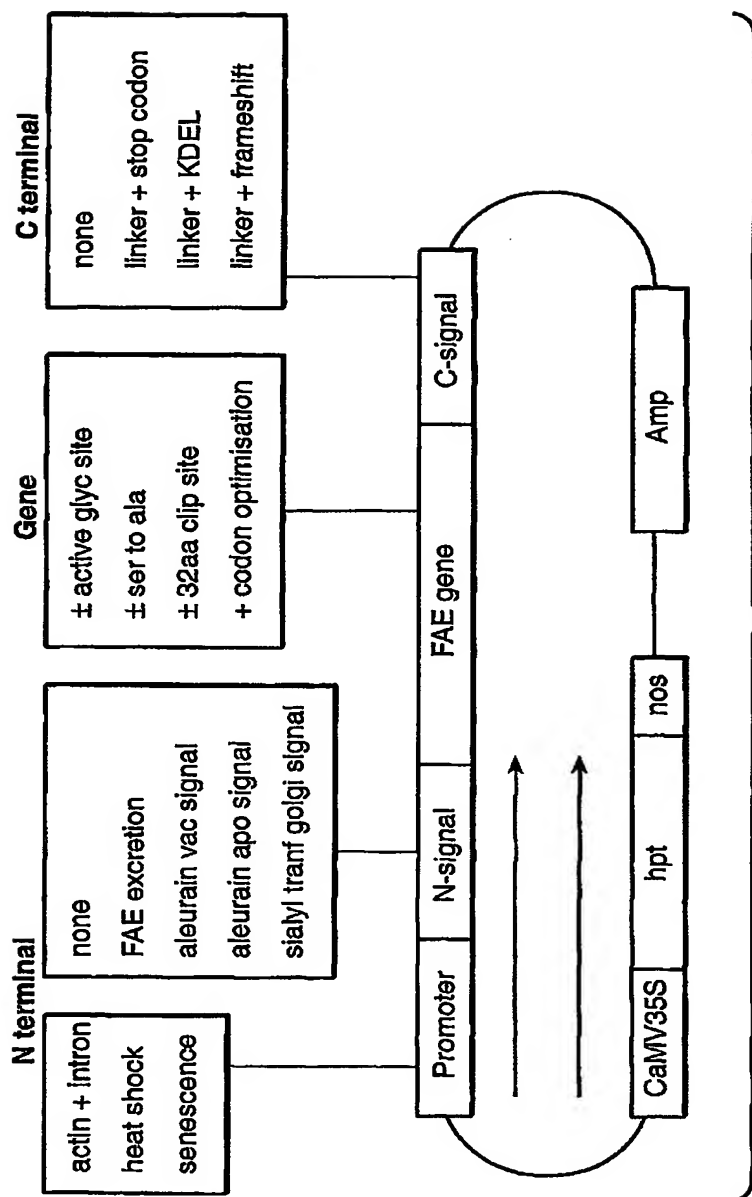
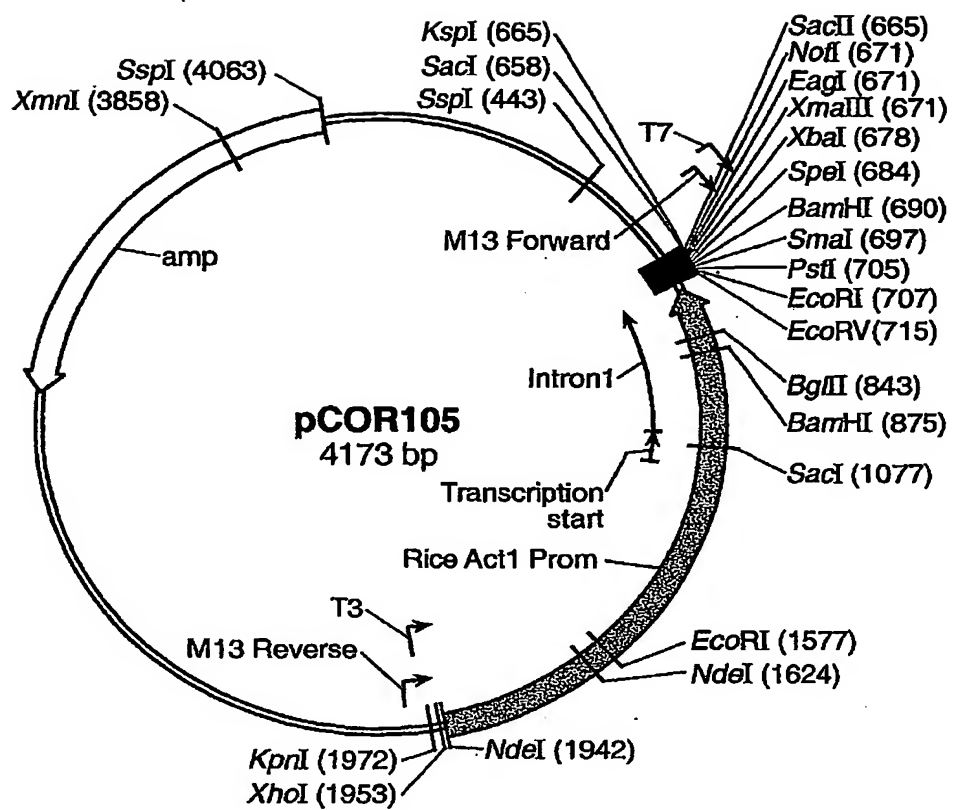
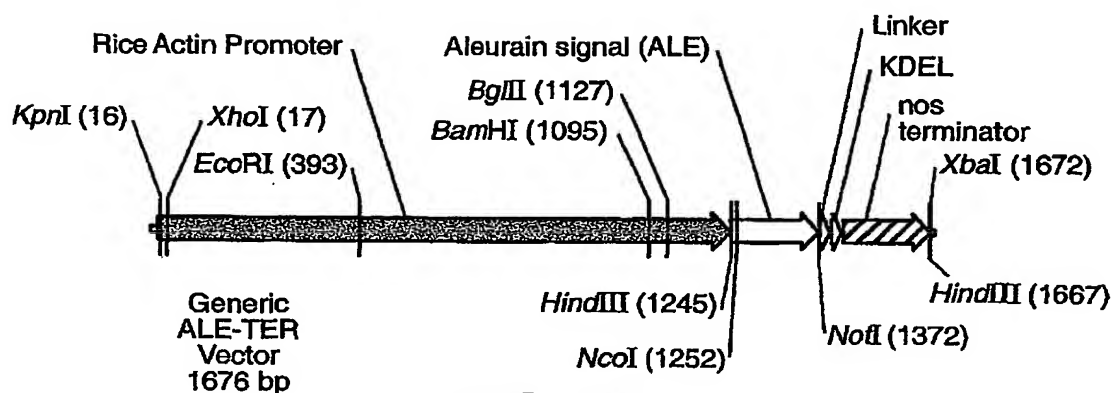


FIG._6

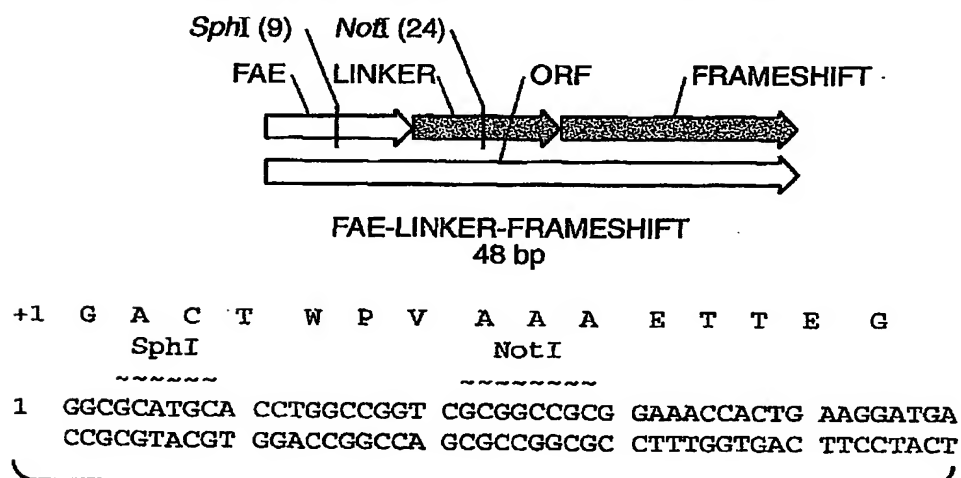
**FIG. 7**

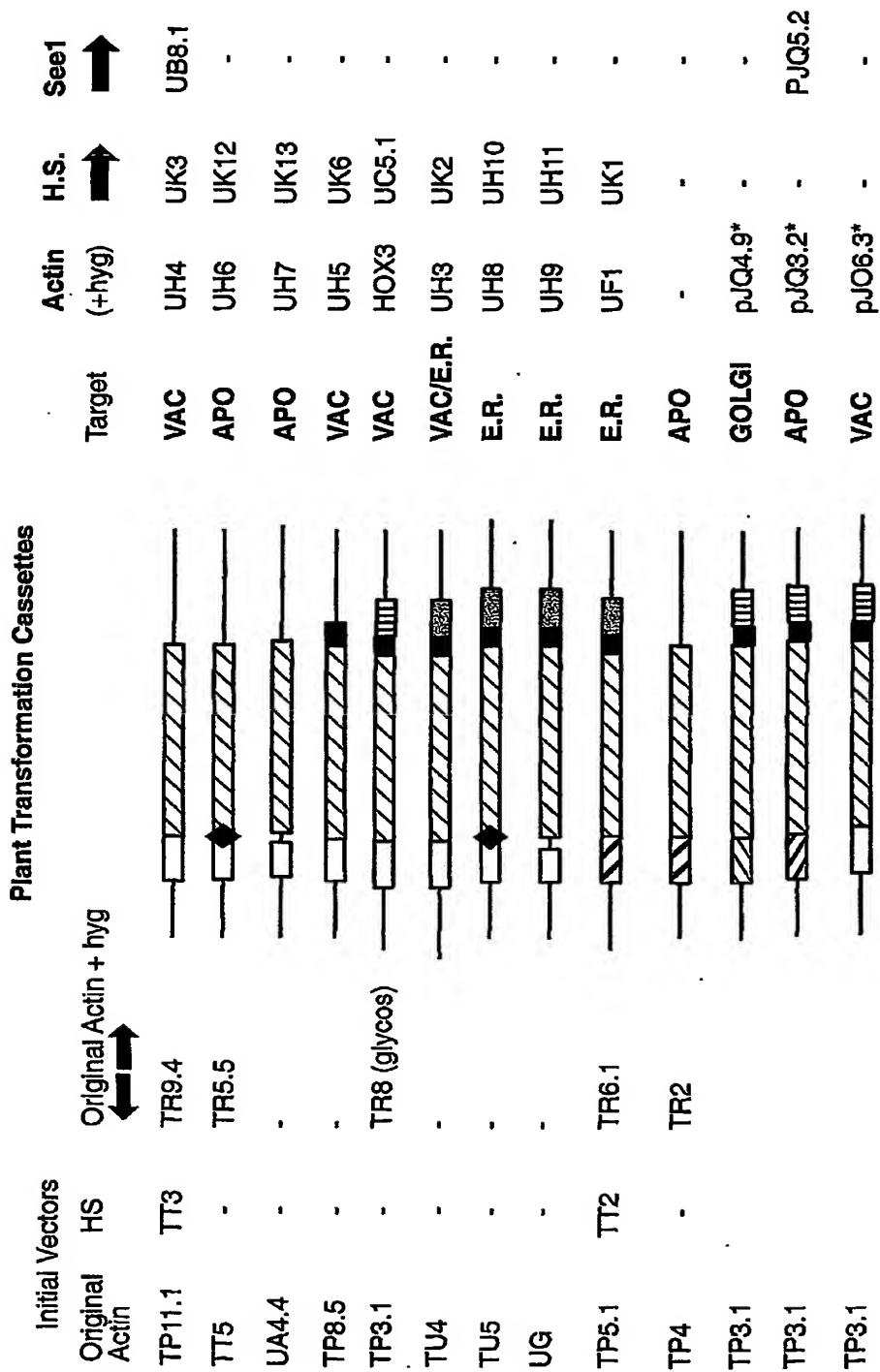
**FIG._8**

KDEL-COOH ER retention sequence

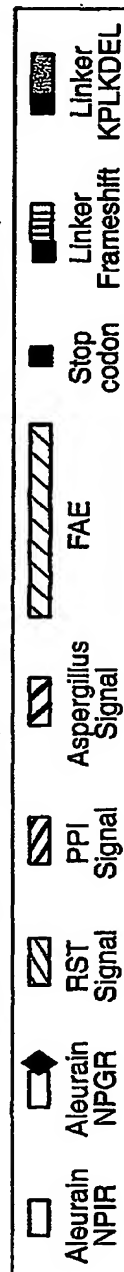
NotI

A A A K P L K D E L *
1 GCGGCCGCGA AACCACTGAA GGATGAGCTG TAA

FIG._9**FAE-LINKER-FRAMESHIFT Structure and Sequence****FIG._10**



* - Modified Actin Promoter (Kpn1-EcoR1 Deletion and Restored NCO Site)

**FIG. 11**

Vectors**Original Actin promoter in pCOR105**

	Target	Signal sequences	Vectors
(i)	APO	- aleurain-NPGR-FAE	pUH6, pTT5, TT5.5, pTT5.1
		- aleurain-delNPIR -FAE	pUH7, pUA4.4,
(ii)	ER	- aleurain-NPGR-FAE-linker-KDEL	pTU5, pUH8,
		- aleurain-delNPIR-FAE-linker-KDEL	pUG4, pUH9,
(iii)	VAC	- aleurain-NPIR-FAE	pTP11.1, pTR9.4, pUH4, pUK3,
(iv)	ER/VAC	- aleurain-NPIR-FAE-linker-KDEL	pTU4, pUH3,
(v)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pUA1K3, pTP3.1, pUC5.11
(vi)	VAC	- aleurain-NPIR-FAE-linker-stop	pTP8.5, pUH5
(vii)	ER	- Aspergillus signal -FAE-KDEL	pTP5.1, pTP6.1, pUF1,

Modified actin promoter (Kpn1-EcoR1 deletion and restored NCO site)

(i)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pJ06.3
(ii)	GOLGI	- RST-FAE-linker-frameshift	pJQ3.2
(iii)	APO	- PPI-FAE-linker-frameshift	pJQ4.9

Heat-shock promoter

(i)	APO	- aleurain-NPGR-FAE	pUH12
		- aleurain-delNPIR-FAE	pUH13
		- Aspergillus signal-FAE	pTP4a2, pTR2.22,
(ii)	ER	- aleurain-NPGR-FAE-linker-KDEL	pUH10
		- aleurain-delNPIR-FAE-linker-KDEL	pUH11
(iii)	VAC	- aleurain-NPIR-FAE	pUK3, pTT3
(iv)	ER/VAC	- aleurain-NPIR-FAE-linker-KDEL	pUK2
(v)	VAC	- aleurain-NPIR-FAE-linker-frameshift	pUC5.11, pHOX3
(vi)	VAC	- aleurain-NPIR-FAE-linker-stop	pUK6
(vii)	ER	- Aspergillus signal -FAE-KDEL	pUK1, pTT2

Senescence promoter

(i)	APO	- See1-PPI-FAE-linker-frameshift	pJQ5.2
(ii)	VAC	- See1-aleurain-deleted NPIR-FAE	pUB8.1

FIG. 12

ALEURAIN-NPIR (Vacuolar) and NPGR (Apoplast) Structure and Sequence

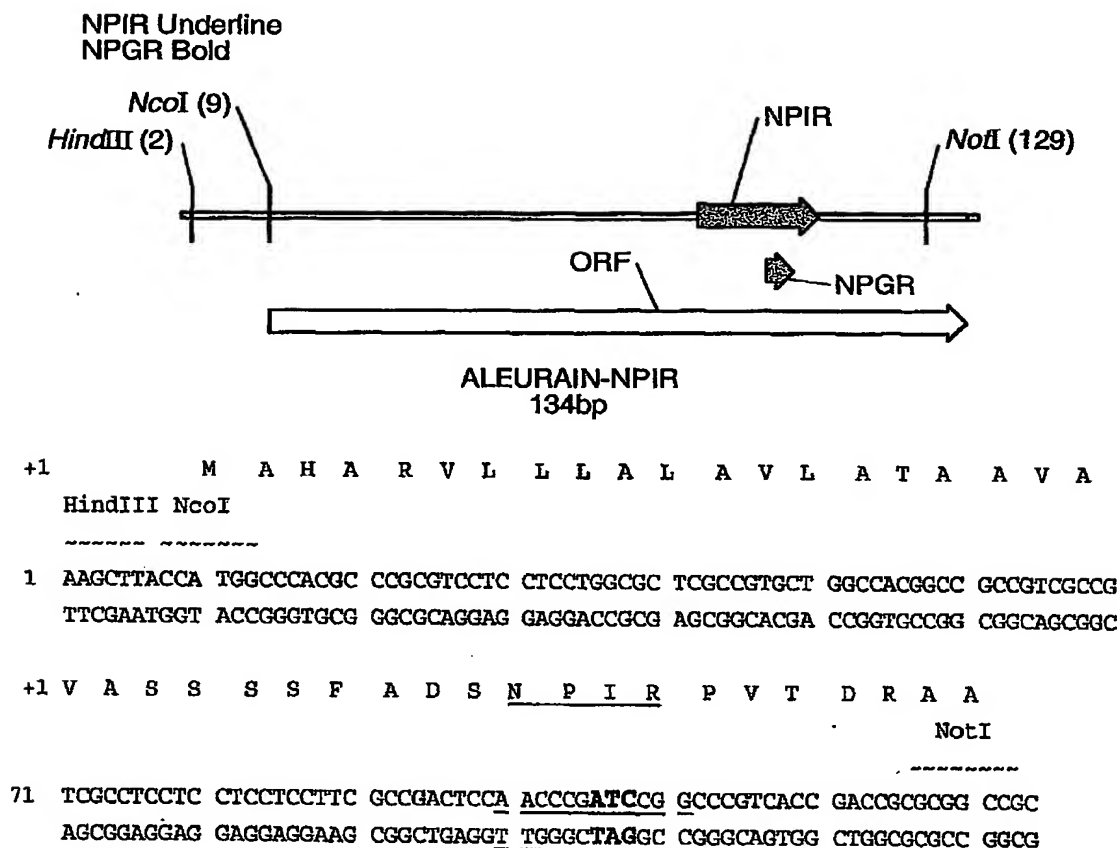


FIG. 13

RAT SIALYL TRANSFERASE Golgi signal sequence

HindIII
 ~~~~~  
 M I H T N L K K K F S L F I L V F L L F A  
 1 AAGCTTACCA TGATCCACAC CAACCTCAA AAGAAGTTCT CCTCTTCAT CCTGGTCTTC CTCTCTCTCG  
 . V I C V W K K G S D Y E A L T L Q A K E F Q M  
 71 CCGTGATCTG CGTGTTGAAG AAGGGCTCCG ACTACGAGGC CCTCACCTTC CAAGCCAAGG AGTTCCAAAT

NotI  
 ~~~~~  
 . A A
 141 GGCGGCGGC

FIG._14

POTATO PROTEASE INHIBITOR II Apoplast signal sequence

HindIII
 ~~~~~  
 M X V H K E V N F V A Y L L I V L G L L L  
 1 AAGCTTACMA TGGMCGTGCA CAAGGAGGTS AACTTCGTSG CCTACCTCCT GATCGTSCCTC  
 GGCCTCCTCT

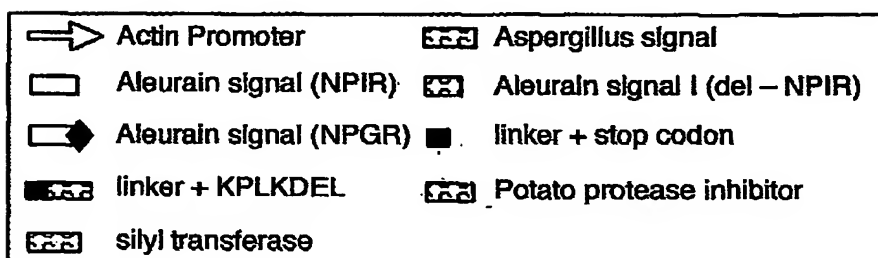
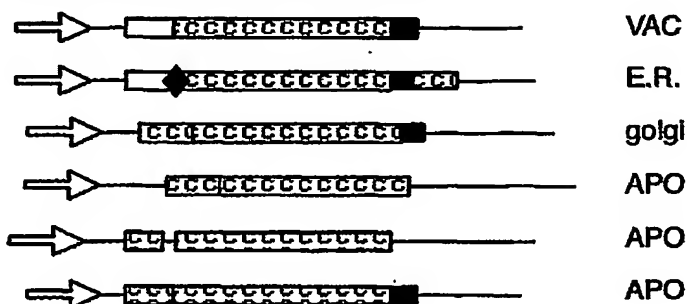
NcoI  
 ~~~~~  
 . L V S A M E H V D A K A C T X E C G N L
 G F G .
 71 TGCTCGTSTC CGCCATGGAG CAGGTGGAGG CCAAGGCCCTG CACCCCKGAG TCGGGCACC
 TCGGCTTCGG

NotI
 ~~~~~  
 . I C P A A A  
 141 CATCTGCCCG GCGGCGGCC

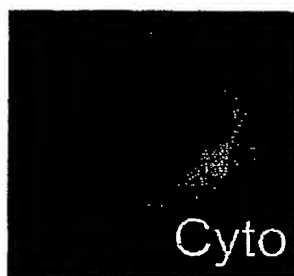
FIG.\_15

### Targeting Expression of gfp to Different Cell Compartments

#### Actin Promoter Targeting Vectors



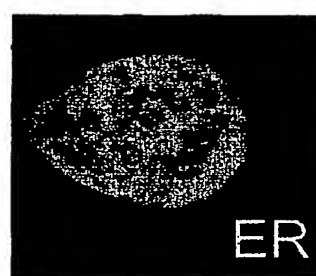
**FIG. 16A**



**FIG. 16B**



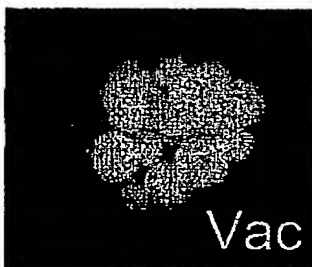
**FIG. 16C**



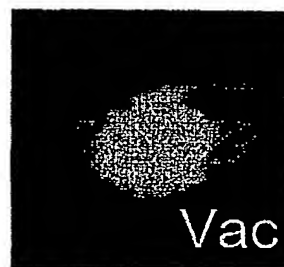
**FIG. 16D**



**FIG. 16E**

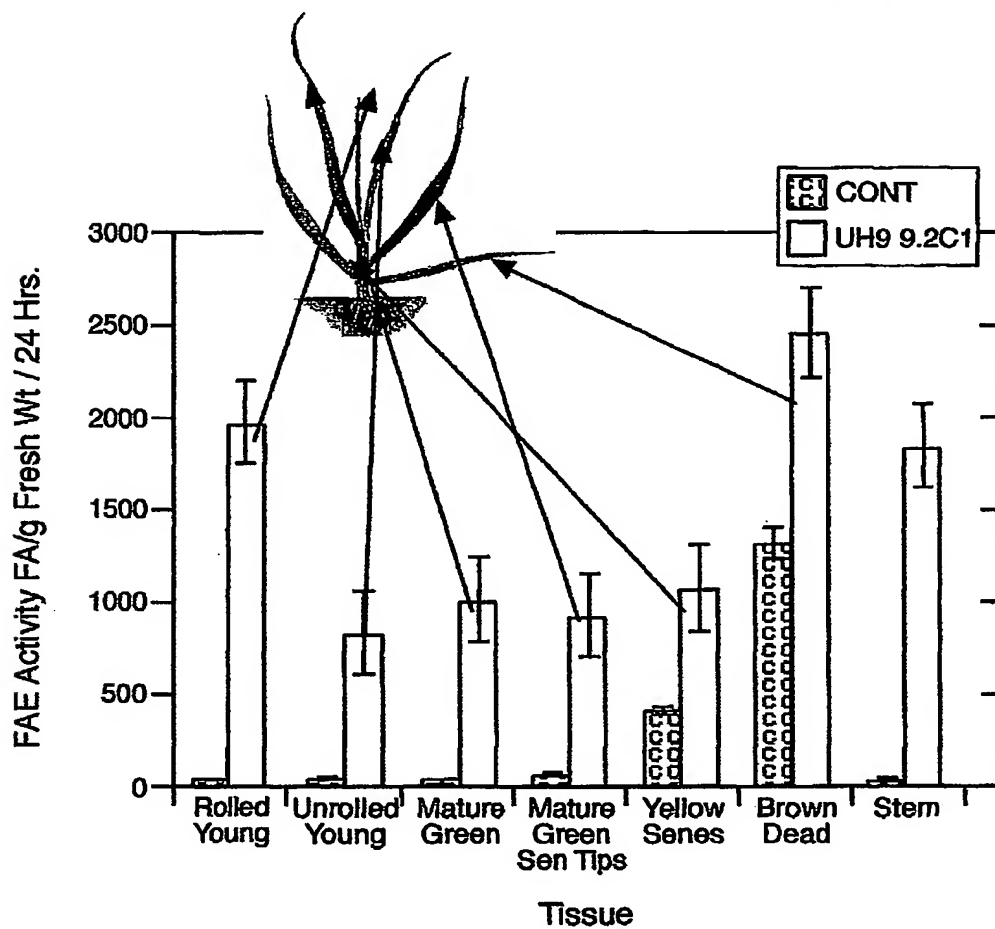


**FIG. 16F**



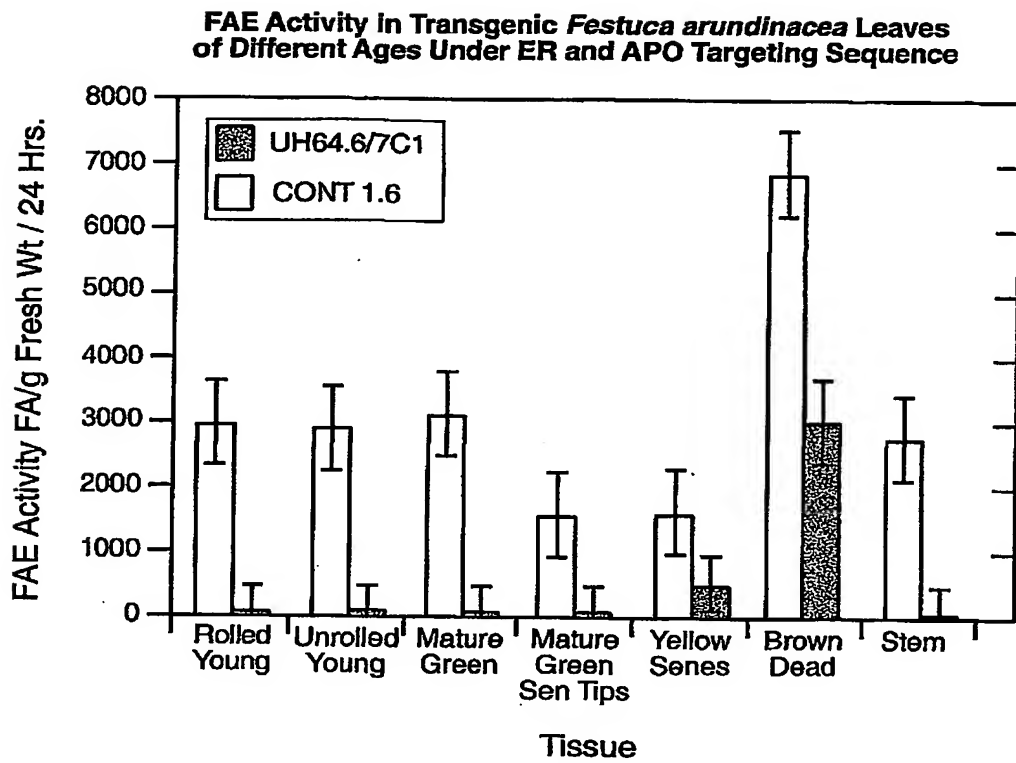
**FIG. 16G**

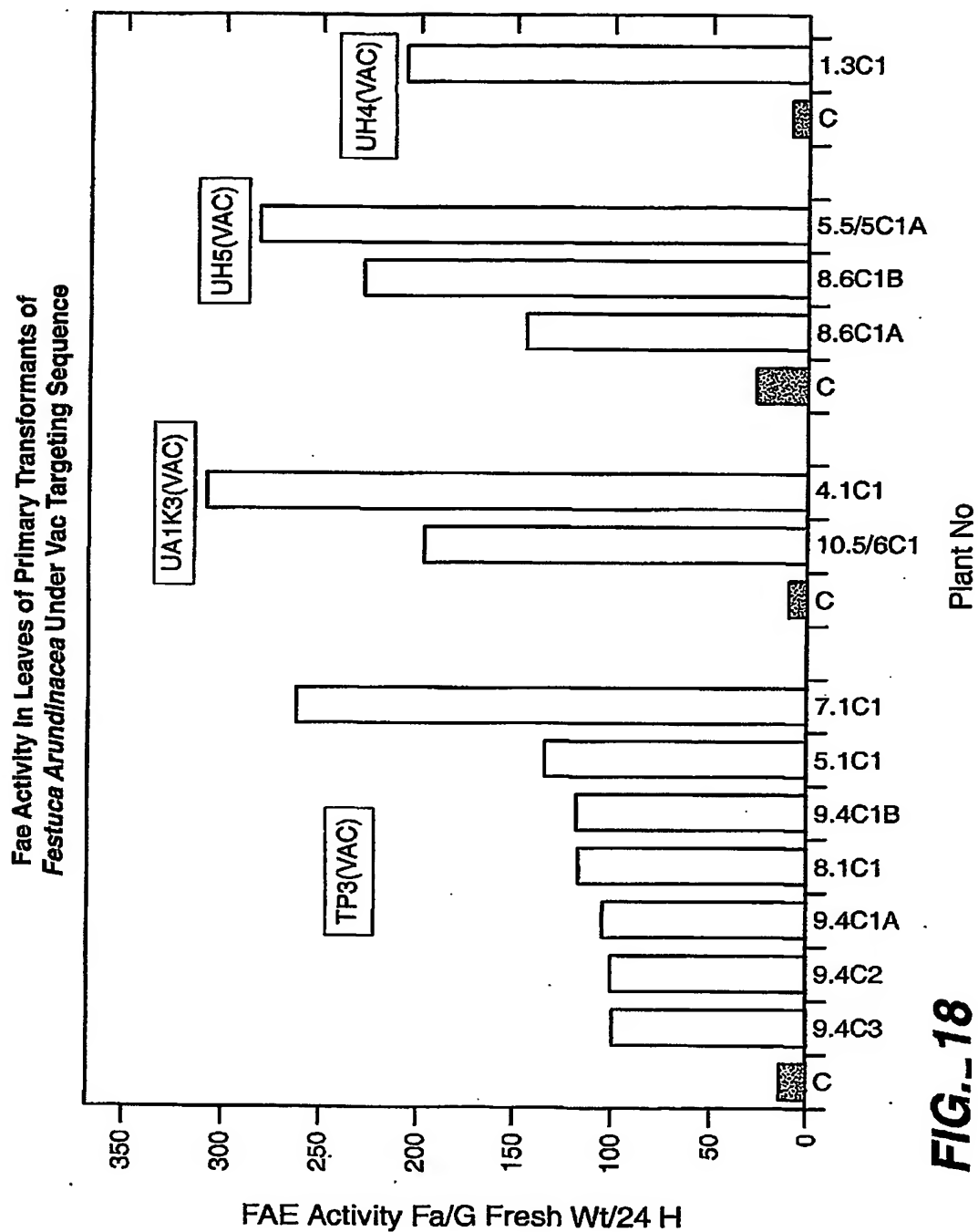
**FAE Activity in Transgenic *Festuca arundinacea* Leaves  
of Different Ages Under ER and APO Targeting Sequence**

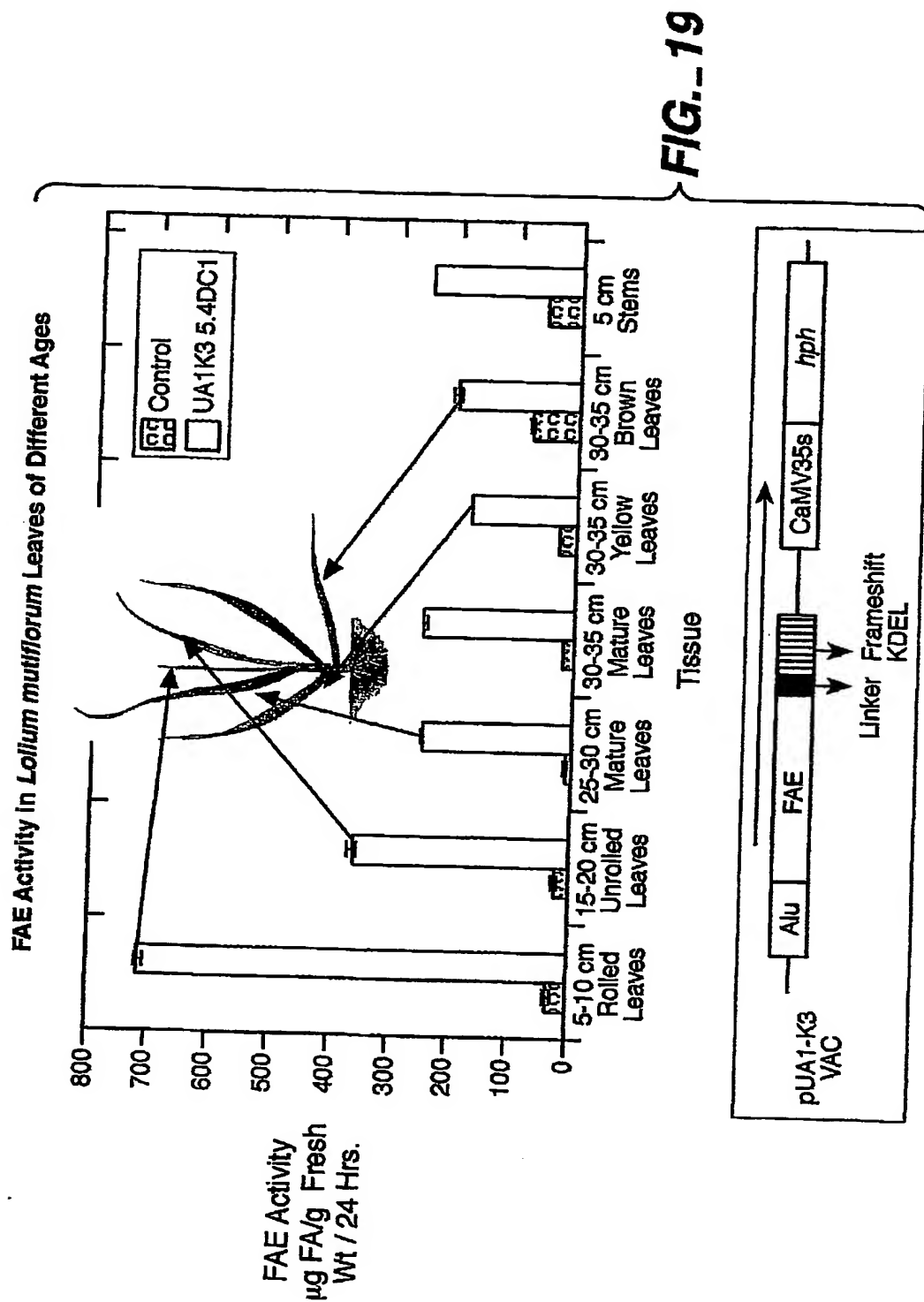


**FIG. 17A**



**FIG. 17B**

**FIG. 18**



FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

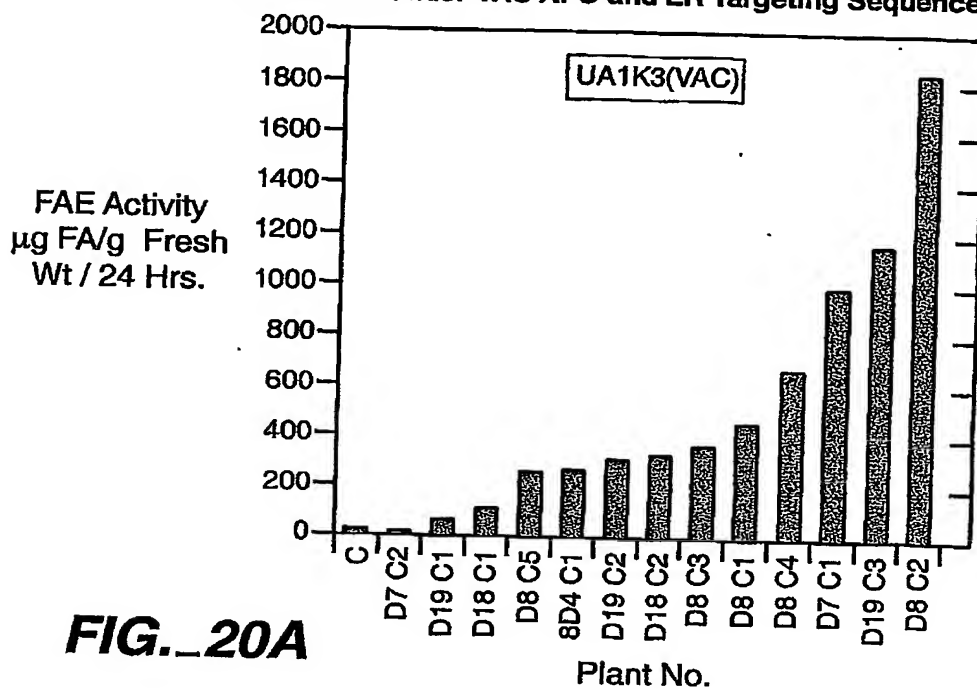


FIG. 20A

FAE Activity in Leaves of Primary Transformants of *Lolium multiflorum* Under VAC APO and ER Targeting Sequence

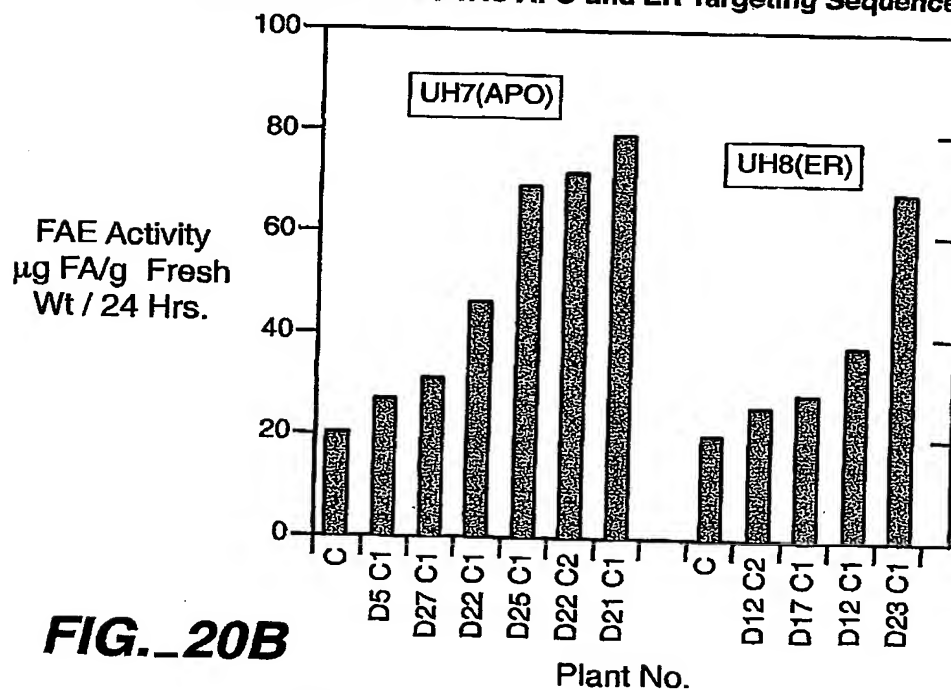


FIG. 20B

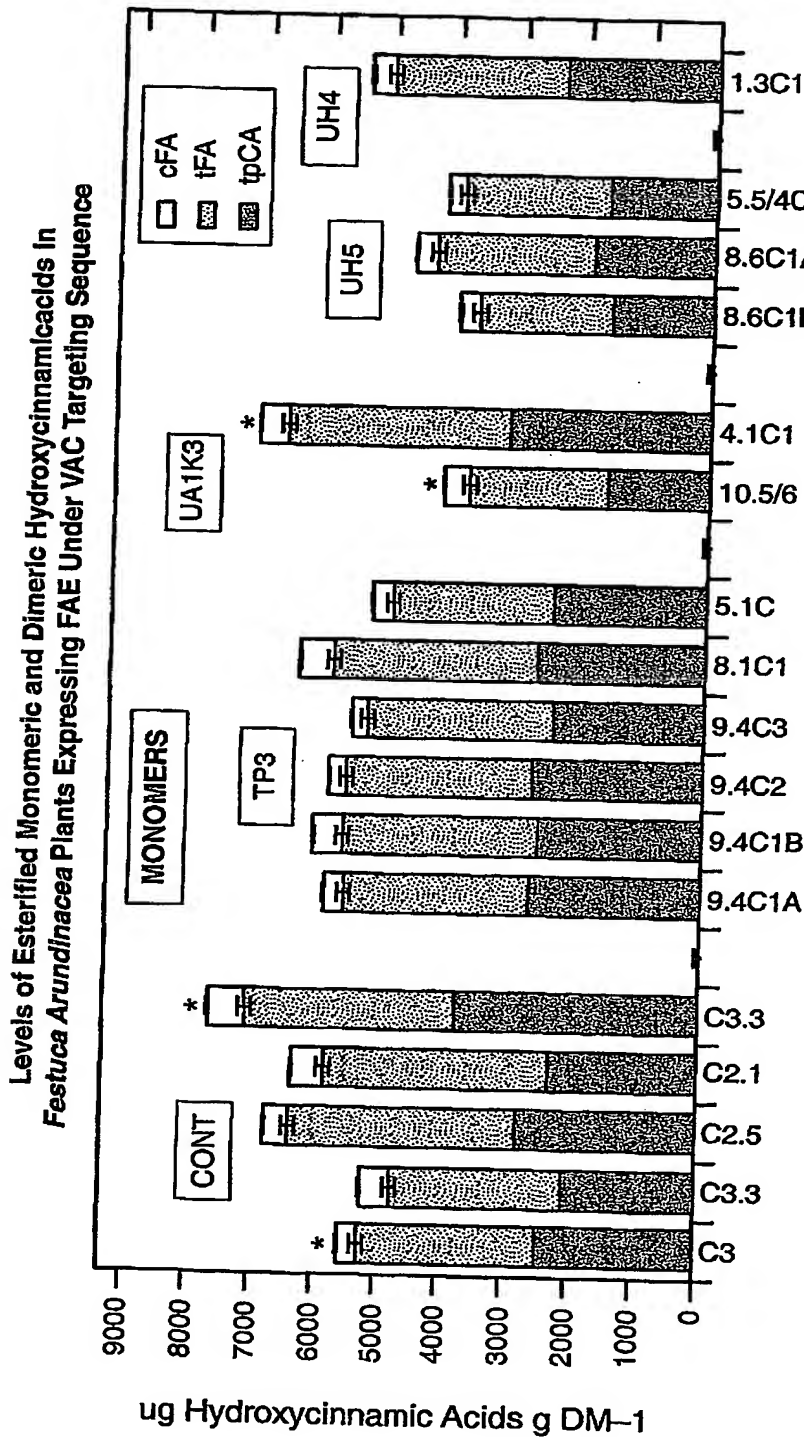


FIG. 21A

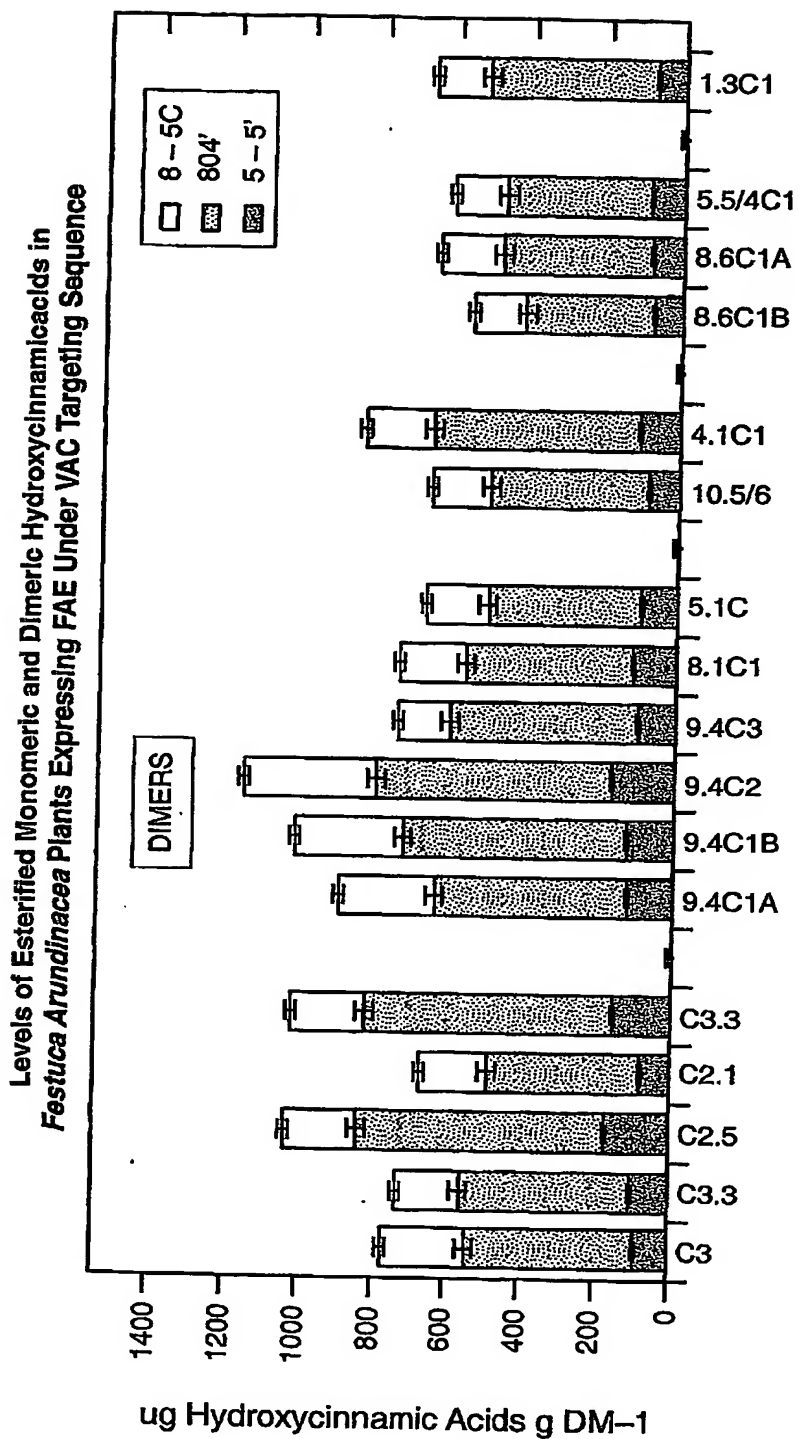


FIG. 21B

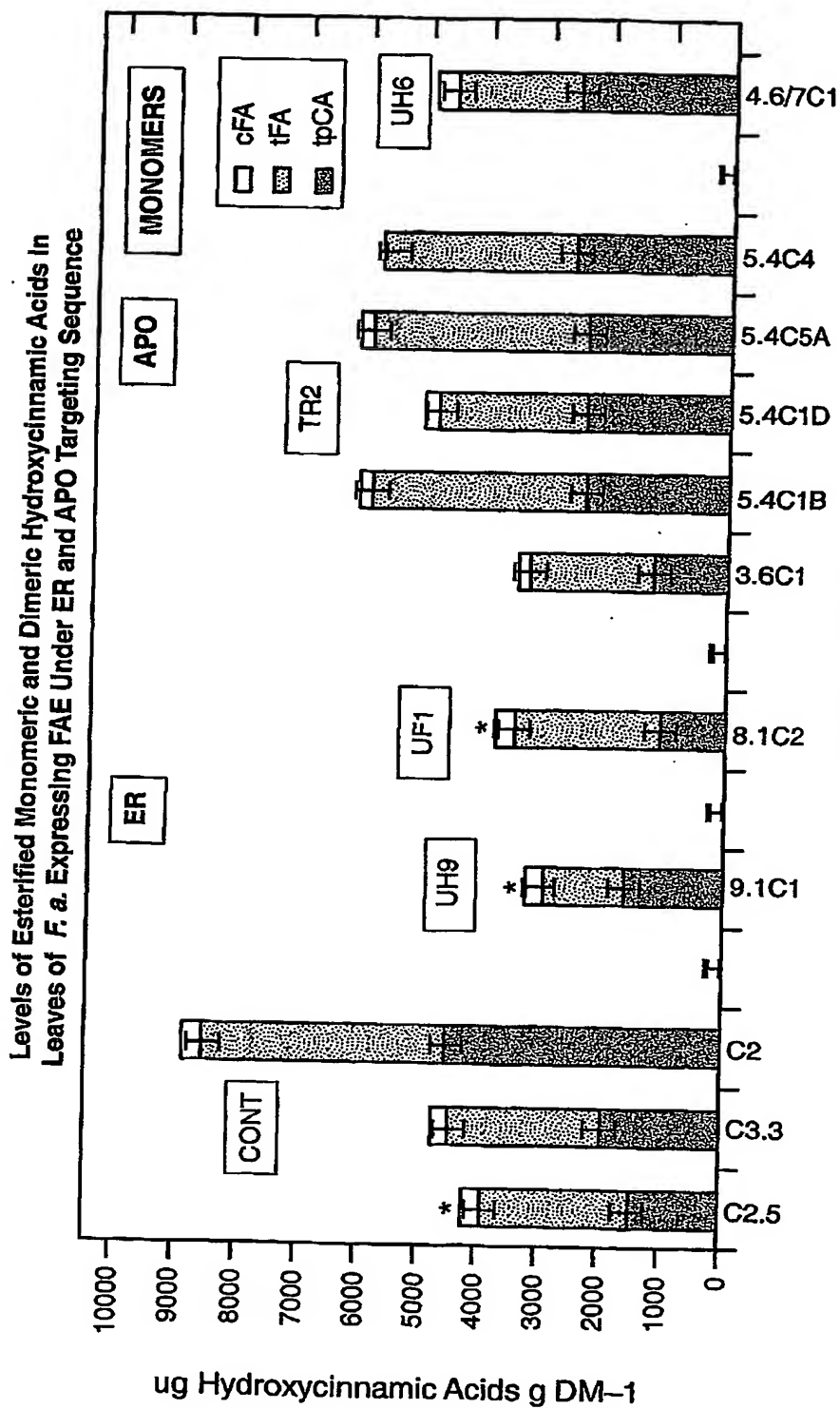
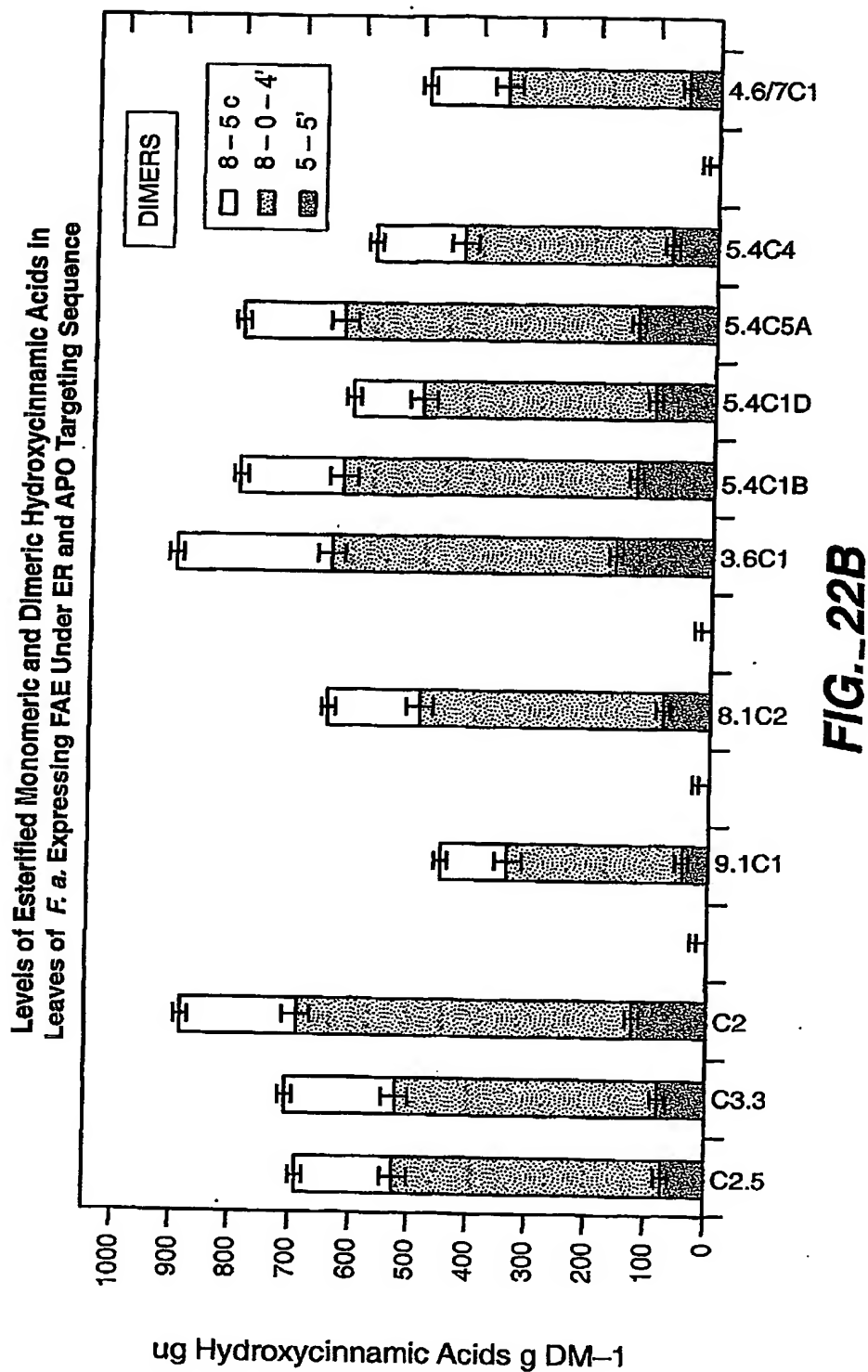


FIG. 22A





In Vitro Dry Matter Digestibility of Leaf Tissue of Mature  
*Festuca arundinacea* Plants Expressing FAE Under an Actin Promoter

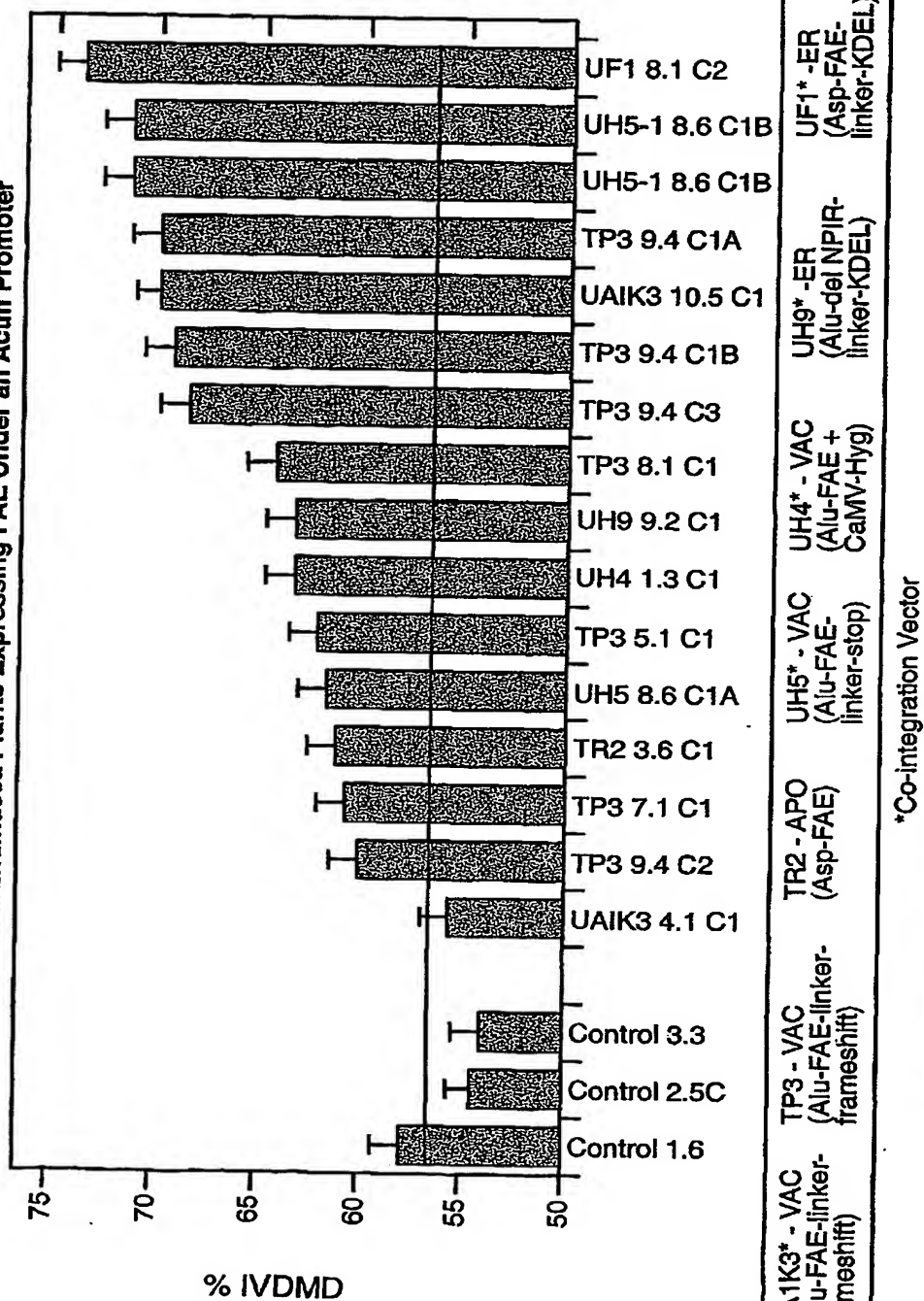
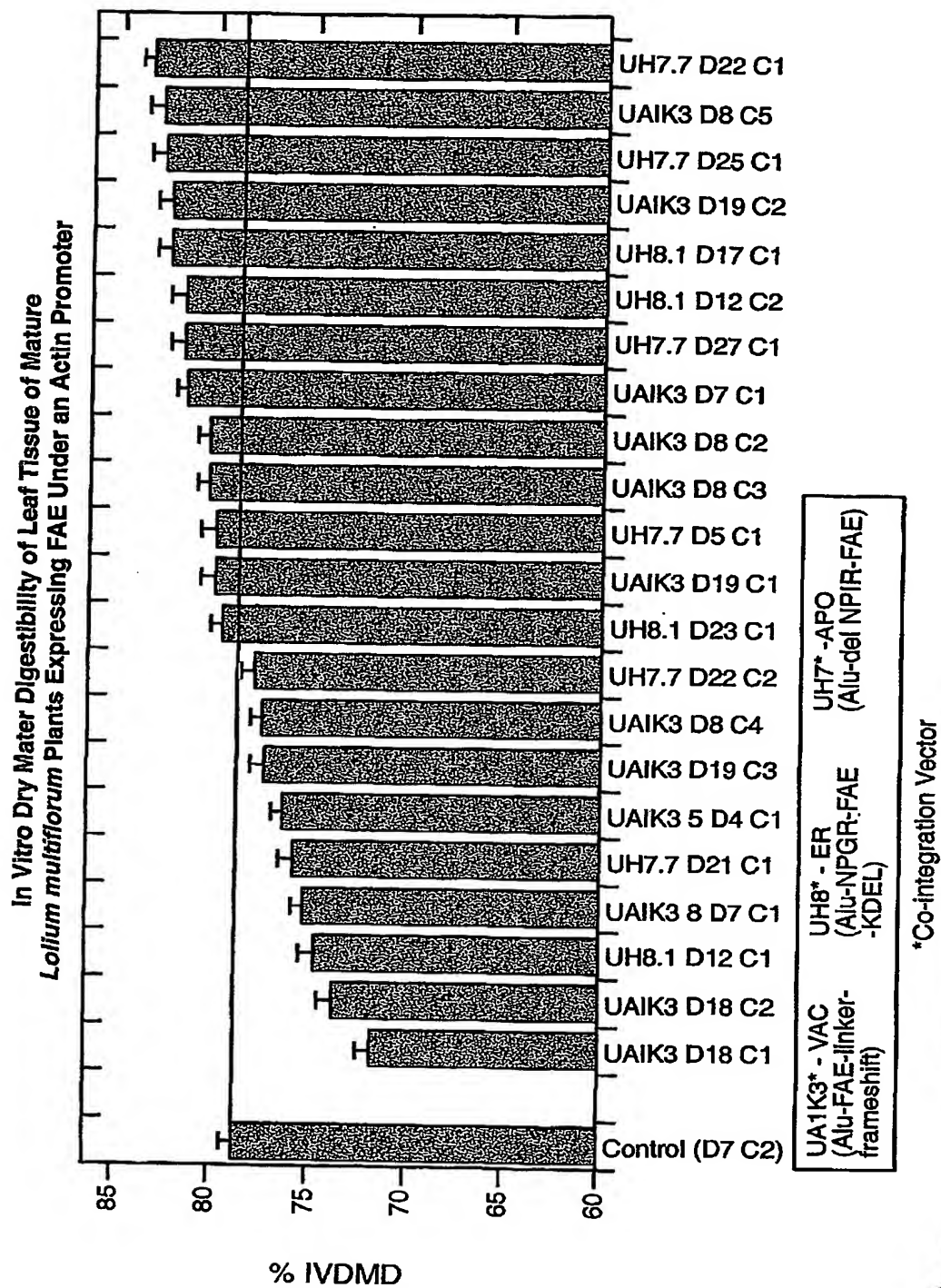


FIG. 23

**FIG. 24**

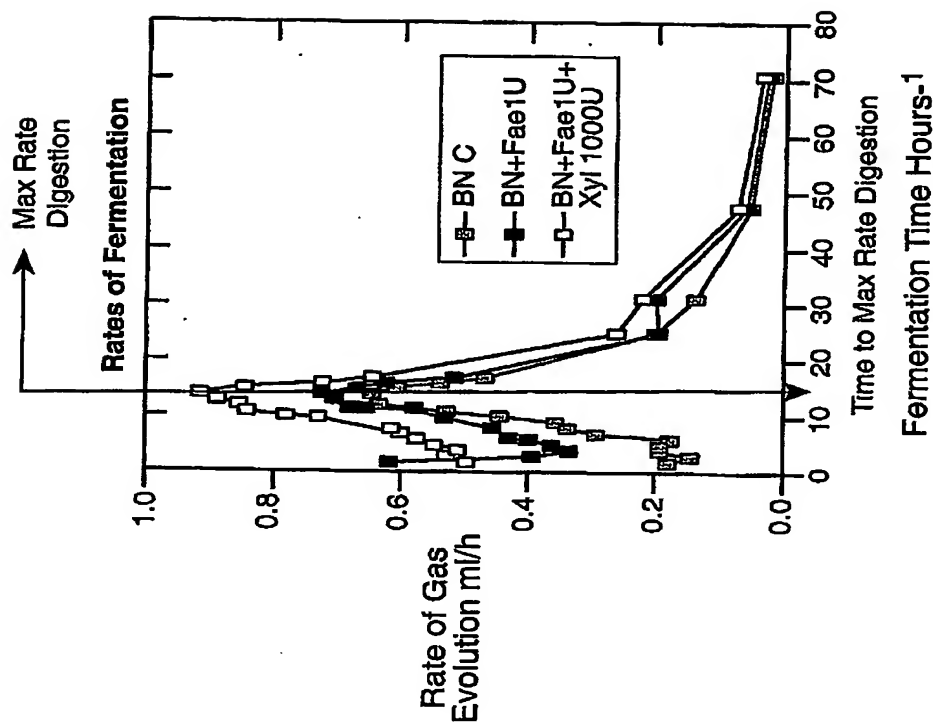


FIG.\_25B

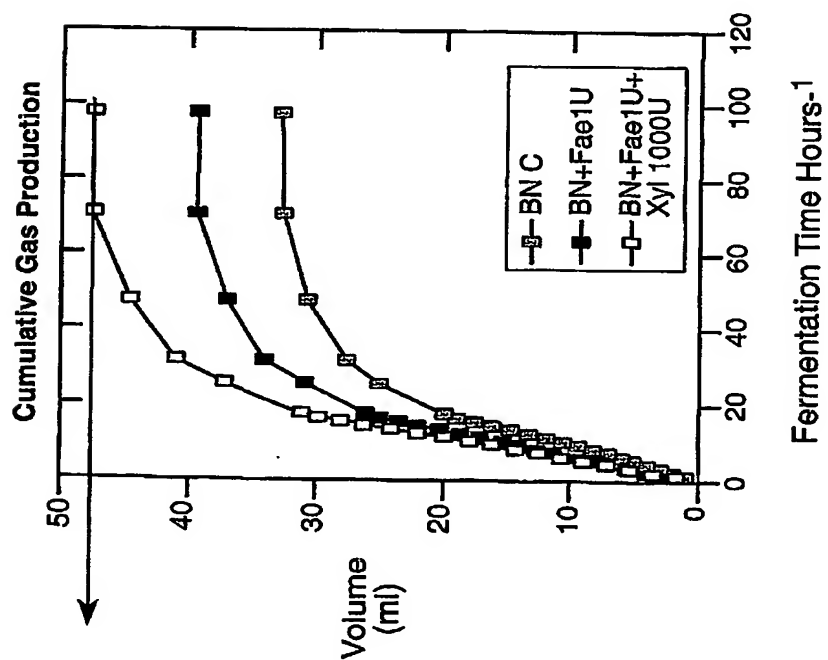
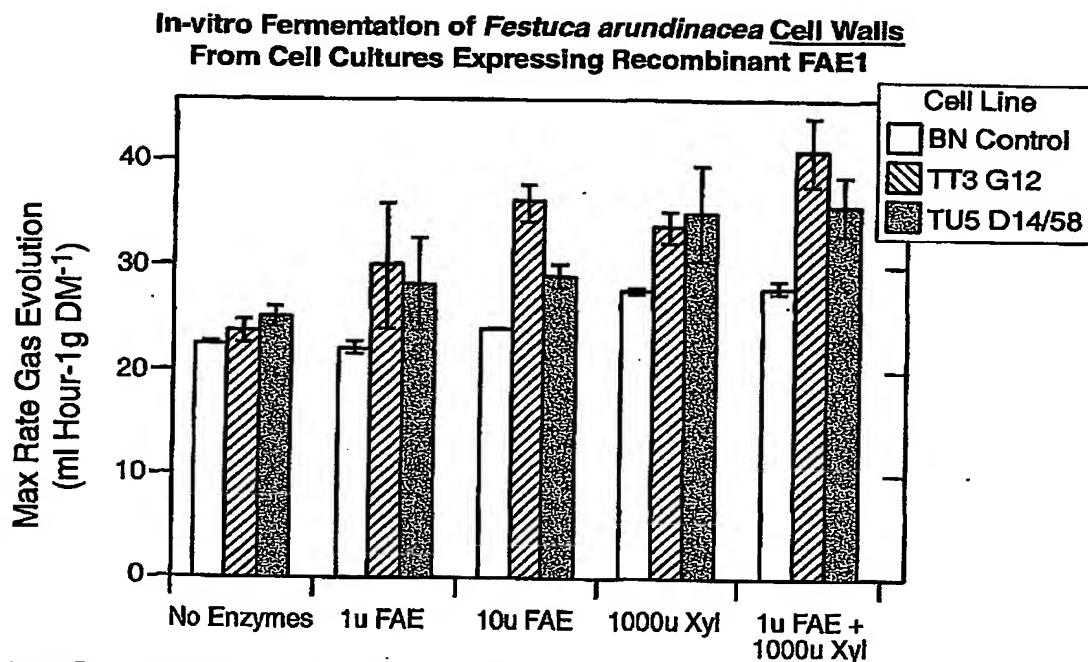
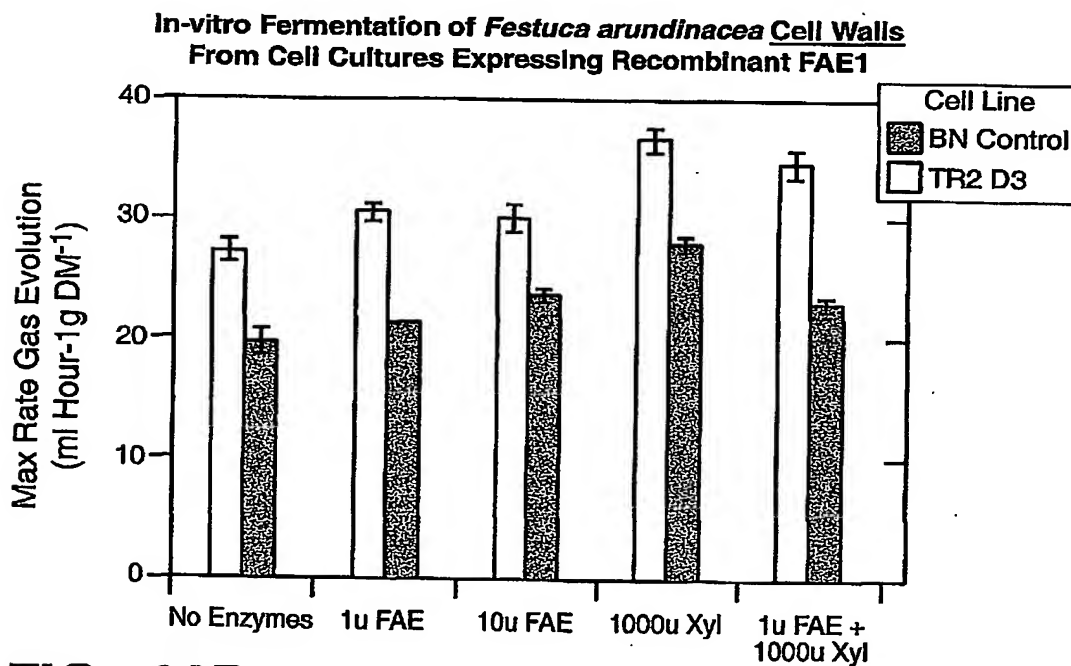


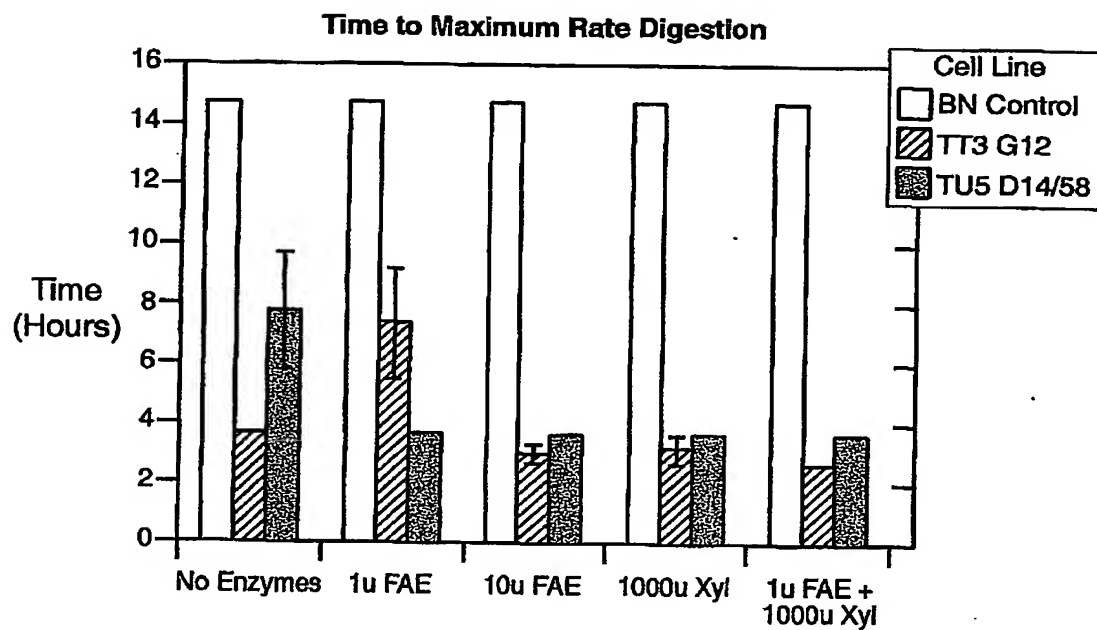
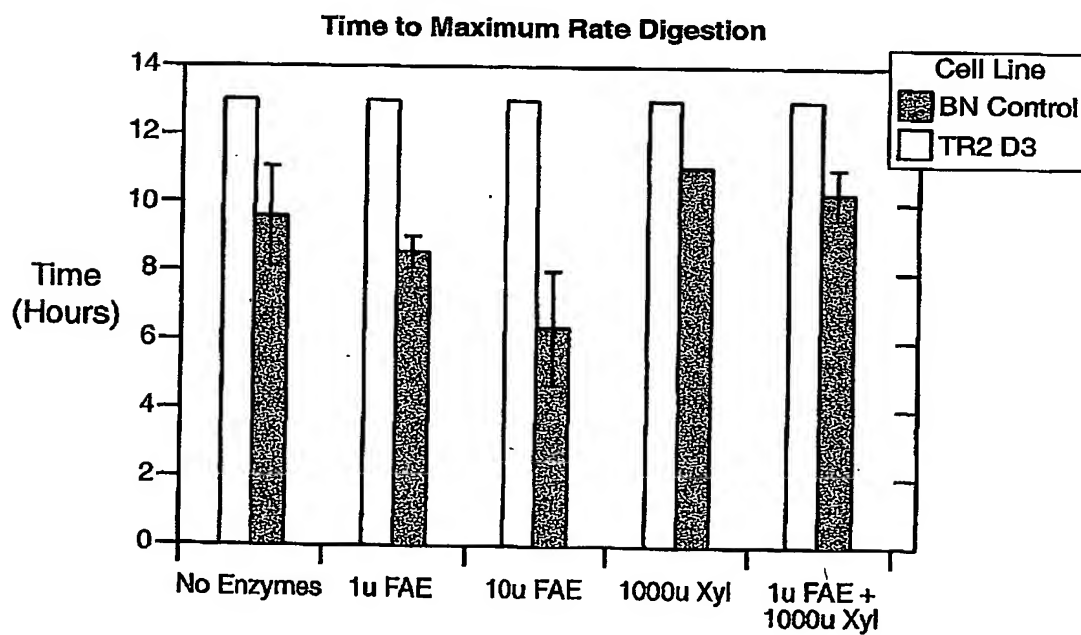
FIG.\_25A

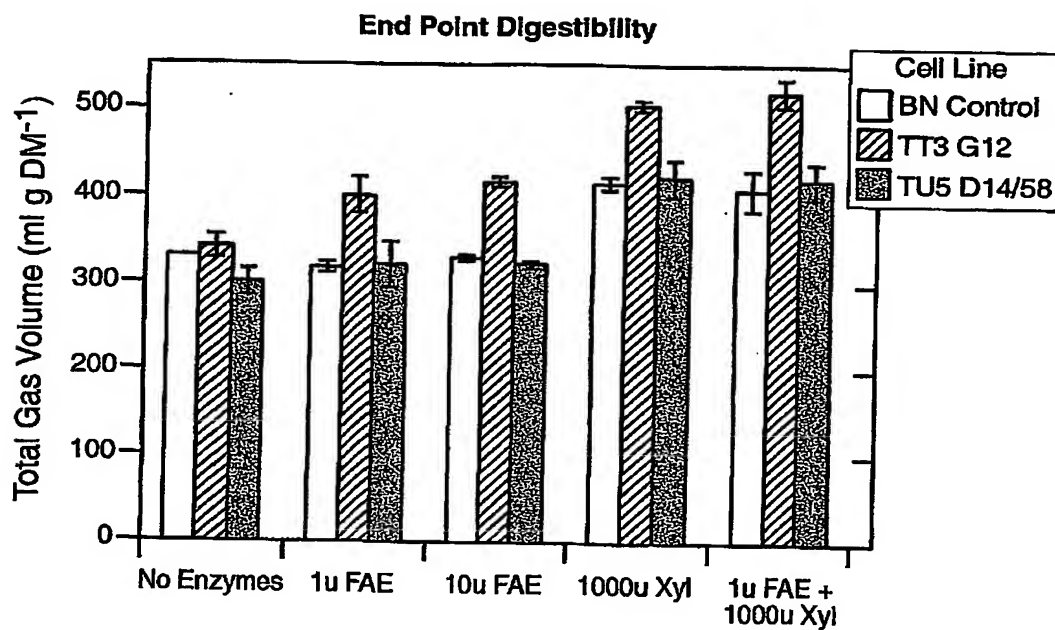
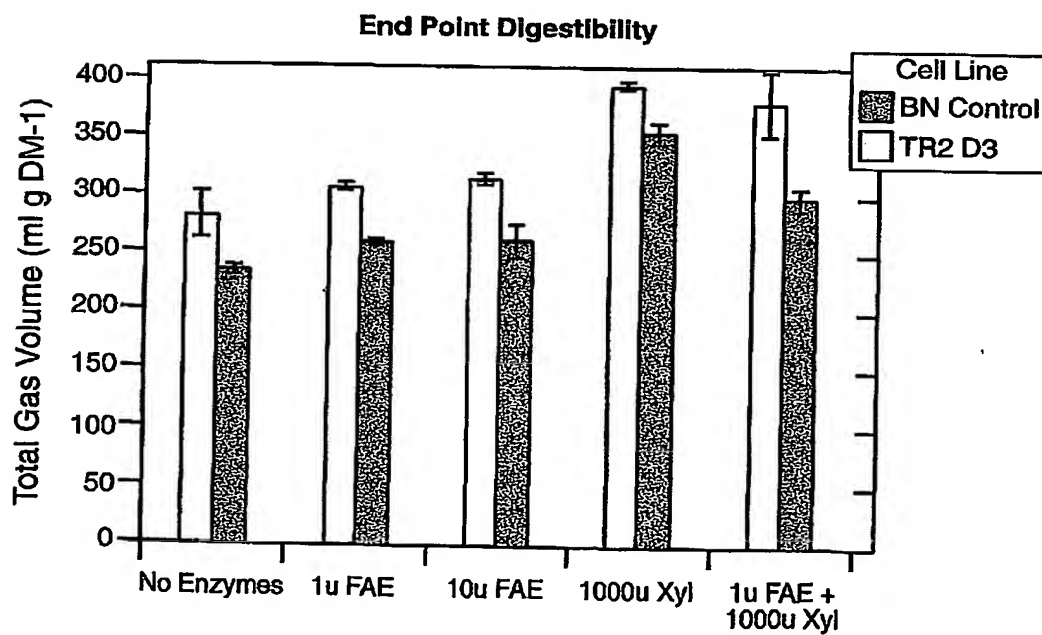


**FIG. 26A** Maximum Rate of Digestion

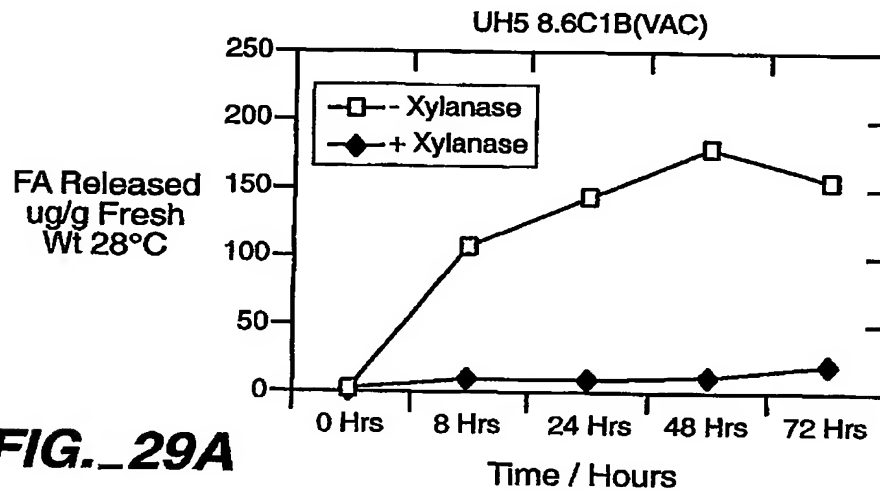


**FIG. 26B** Maximum Rate of Digestion

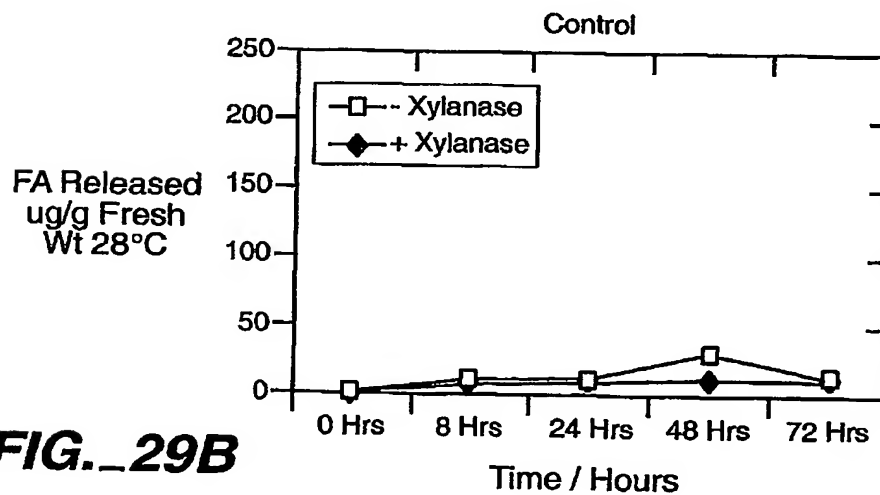
**FIG. 27A****FIG. 27B**

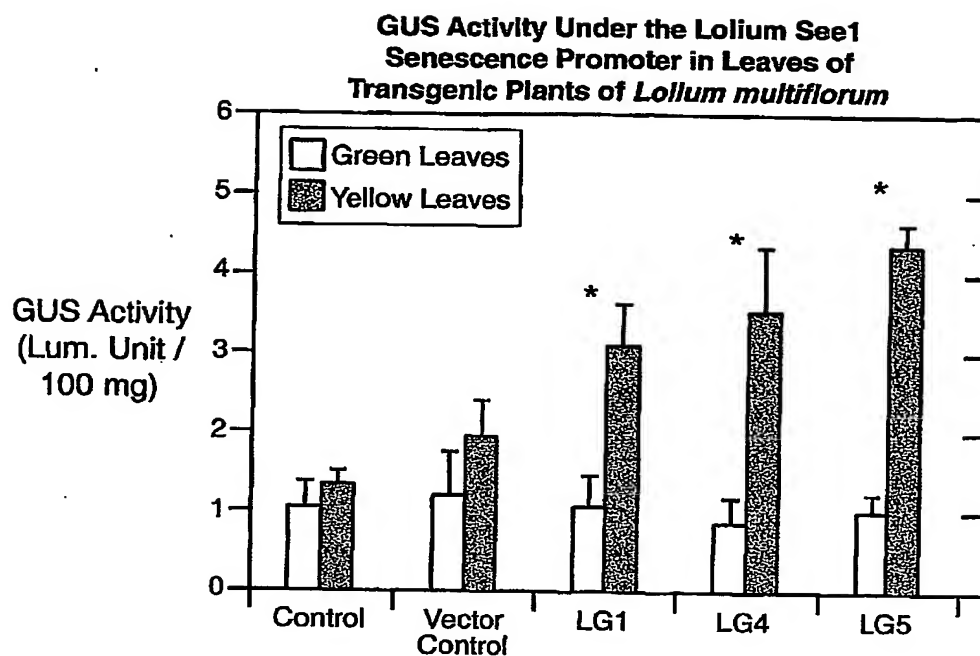
**FIG.\_28A****FIG.\_28B**

**Kinetics of FAE Activity by Ferulic Acid Release  
from Cell Wall under Self Digestion in *Festuca arundinacea*  
and Stimulation by Xylanase**

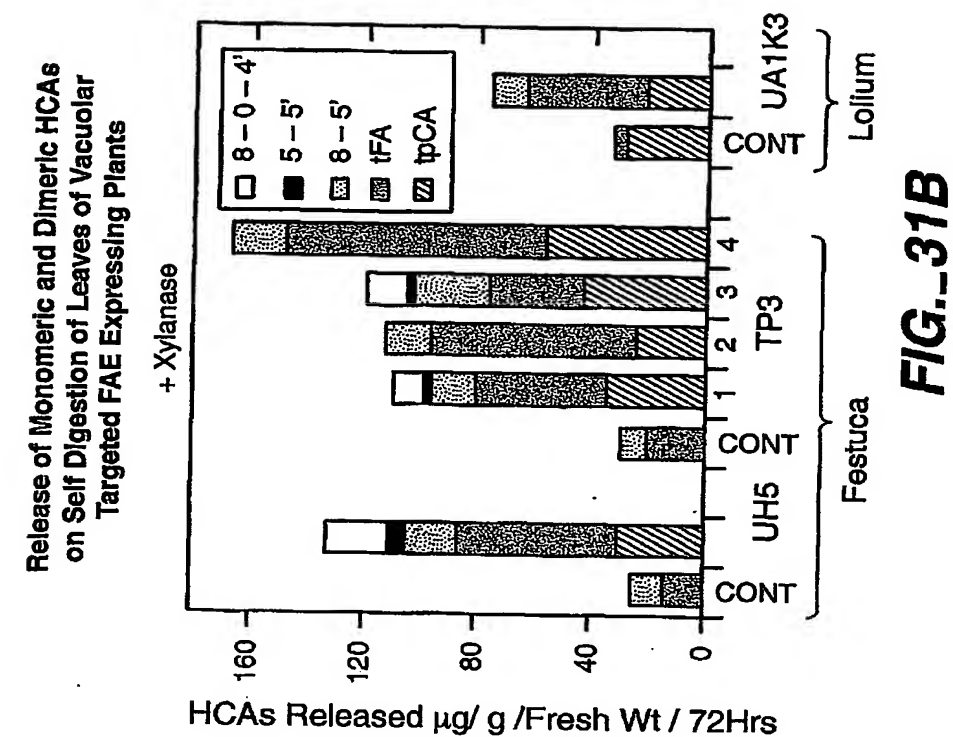
**FIG.\_29A**

**Kinetics of FAE Activity by Ferulic Acid Release  
from Cell Wall under Self Digestion in *Festuca arundinacea*  
and Stimulation by Xylanase.**

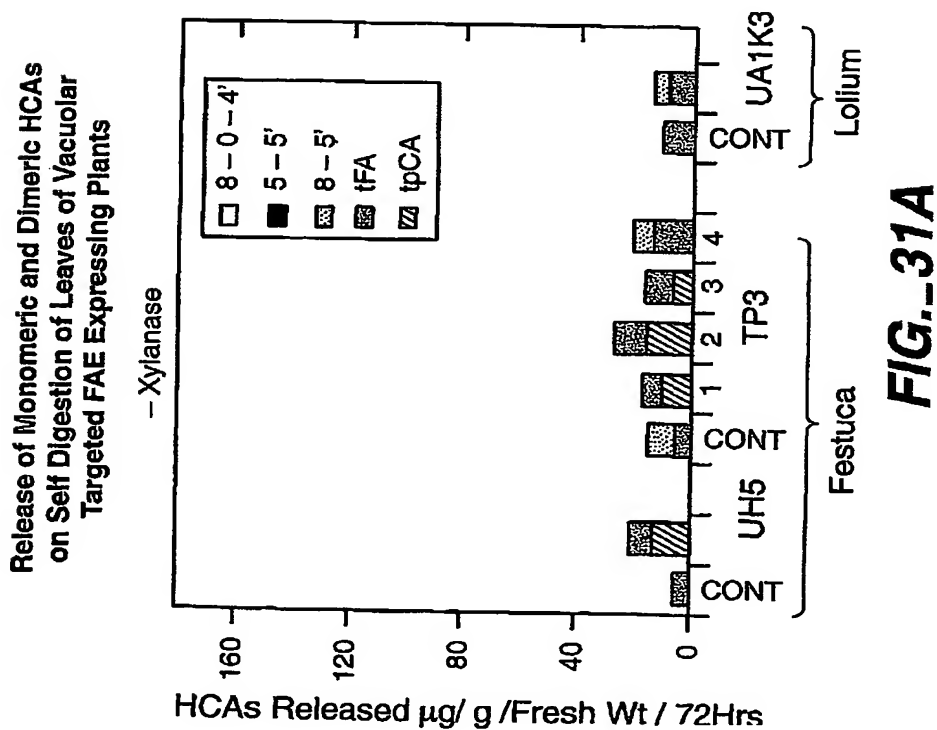
**FIG.\_29B**

**FIG.\_30**

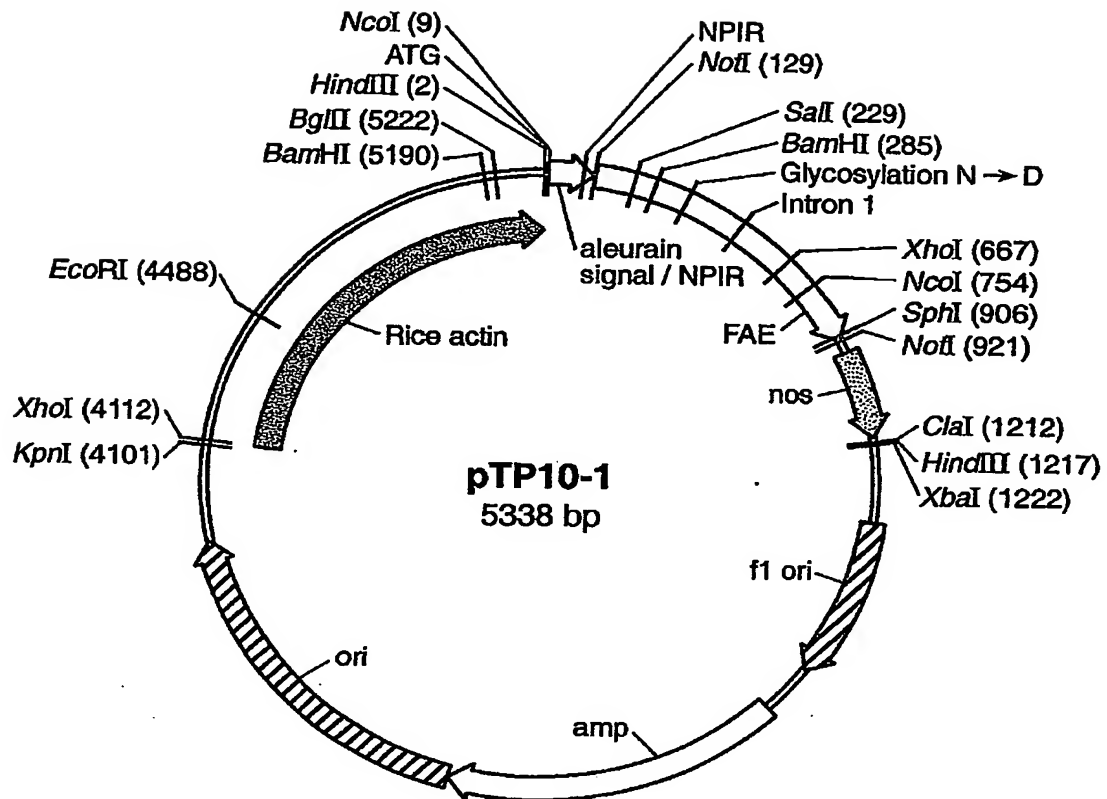




**FIG.\_31B**



**FIG.\_31A**

**FIG.\_32A**

```

                                NotI
                                ~~~~~
HindIII
~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGGTTCCT CTCTGGCGC TCCCGTGCT GGCCACGGC GCCGTGCGG

                                NPIR
                                ~~~~~
 . A S S S F A D S N P I R P V T D R A A A S T .
71 TGGCTCCTC CTCCTCCTC GCCGACTCA ACCGATCCG GCCGTCAAC GACCGCGG CCGCTCCAC
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGCATC TCCGAAGACC TCTACAGCG TTAGTCGAA ATGGCCACTA TCTCCCAAG TGCCTACGCC

 SalI
                                ~~~~~
      D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGAGA AATTTCACAA TTCTCAACT GACATTACG

                                BamHI
                                ~~~~~
 . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGGACGAC AGCAGCAAAG AAATAATCAC CGTCTCCGT GGCACGTGTA GTGATACGAA

 Glycosylation
                                ~~~~~
      . L Q L D T D Y T L T P F D T L P Q C N G C E V
351 TCTACAATC GATACTGACT ACACCTCAC GCCTTTCGAC ACCCTACAC AATGCAACGG TTGTGAAGTA
  H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGATC GCTTGTCAA CAGCAGGTTA
  . Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GACTACGCG CTGACCGTGA CCGGCCACKC CCTCGCGGC TCCCTGGCG CACTCCTGC
  . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCCAAGCTG TCTGGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGG AACCGCGCAG CGGCAATCAG

```

FIG. 32B

XhoI  
 631 GCCTTGGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA  
 NcoI  
 701 CTGATGCCAA CGACGGCATC CCAAGCTGC CCCGGTGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACGGGG ATGAGTGA GTGCTGTGAG  
 SphI  
 841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGGGC GCATGCACCT  
 NotI  
 911 GGCCGGTTCG GCGCGCGAA ACCACTGAAG GATGAGTGT AAAGAAGCAG ATCGTTCAAA CATTGGCAA  
 981 TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGGCA TGATTATCAT ATAAATTCAG TTGAATTACG  
 1051 TTAAGCAGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATAGA TTAGAGTCCC  
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAACT AGGATAAATT ATCGCGCGCG  
 HindIII  
 1191 GTGTCACTTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTCGC CCTATAGTGA  
 1261 GTCGTATTAC GCGCGTAC TGGCCGTCTG TTTACAACGT CGTACTGGG AAAACCCCTGG CGTTACCCAA  
 1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGTGGC GTAAATAGCA AGAGGCCCGC ACCGATCGCC  
 1401 CTTCCCAACA GTTGGCAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTAA GCGCGCGCGG

FIG. 32C

1471 TGTGGTGGTT ACGCGCAGCG TGACCGGTAC ACTTGCCAGC GCCCTAGCGC CCGCTCCTTT CGCTTCTTTC  
 1541 CTTTCTTTTC TCGCCACGTT CCGCGGCTTT CCCCCTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
 1611 GATTAGTGC TTACCGGCAC TTACCGGCAC AAAAACAATGA TTAGGGTGTAT GGTTCACGTA GTGGGCCATC  
 1681 GCGCTGATAG ACGGTTTTC GCGCTTGGAC TATCTCGGTC TATTTGAGTCC ACGTCTCTTA ATAGTGGACT CTTGTTCCAA  
 1751 ACTGGAACAA CACTCAACC CACTCAACC AATTAACAA AATTTAAC GATTTGCGG GATTTGCGCT ATTTGCGCCT  
 1821 ATTGTTAAA AATGAGCTG AATTAACAA AATTTAAC GATTTGCGG GATTTGCGG GATTTGCGCT ATTTGCGCCT  
 1891 TTAGGTGGCA CTTTTCGGGG CATGAGACAA TAACCTGTAT AAATGCTTCA ATATATATGA AAAAGGAAGA GTATGAGTAT  
 1961 TGTATCCGCT TCAACATTTT CCGTTCGGCC TATTTCCCTT TATTTCCGCA CAGTTGGGTG TTTTGGCTTC CTGTTTTCG  
 2031 TCAACATTTT CCGTTCGGCC TATTTCCCTT TATTTCCGCA CAGTTGGGTG TTTTGGCTTC CTGTTTTCG CTGATCTCA  
 2101 ACGCTGGTGA AAGTAAAGA GATCTTGAAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTTCT  
 2171 ACAGCGGTAA GATCTTGAAG AGTTTTCGCC CCGAAGAACG TTTTCCAATG ATGAGCACTT TTAAGTTCT TTAAGTTCT  
 2241 GCTATGTGGC GCGGTATAT CCGGTATGA CTCACAGTC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAAGAGAT  
 2311 CAGAATGACT TGGTTGAGTA CTCACAGTC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAAAGAGAT TCGGAGGACC  
 2381 TATGCACTGC TGCATTAACC ACCGCTTTT TGCACAACAT GGGGGATCAT GACCTTACTT CAACTTACTT TGAATCGTTG  
 2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGGATCAT GACCTTACTT CAACTTACTT TGAATCGTTG GGAACCGGAG  
 2521 CTGAATGAAG CCATACCAA CCGCTTTT TGCACAACAT GGGGGATCAT GACCTTACTT CAACTTACTT TGAATCGTTG GGAACCGGAG  
 2591 AACTATTAAC TGGCGAATA CCGCTTTT TGCACAACAT GGGGGATCAT GACCTTACTT CAACTTACTT TGAATCGTTG GGAACCGGAG  
 2661 AGTTGCAAGA CCACTTCTGC CCGCTTTT TGCACAACAT GGGGGATCAT GACCTTACTT CAACTTACTT TGAATCGTTG GGAACCGGAG  
 2731 GAGCGTGGGT CTCGCGGTAT CCGCTTTT TGCACAACAT GGGGGATCAT GACCTTACTT CAACTTACTT TGAATCGTTG GGAACCGGAG  
 2801 ACACGACGGG GAGTCAGGA ACTATGATG AACGAATAG ACAGATCGCT CTTAGATTTG ATTTAAACTT CCTCAGTGT  
 2871 TAAGCATGG TAACTGTGAG ACCAAGTTT GATAATCTCA TGACCAAAAT CTTAGATTTG ATTTAAACTT CCTCAGTGT  
 2941 TTTAAAGGA TCTAGGTGAA GATCCTTTT GATAATCTCA TGACCAAAAT CTTAGATTTG ATTTAAACTT CCTCAGTGT  
 3011 TCCACTGAGC GTCAGACCC GTAGAAAAGA TCAAAGGATC TTTTGTGAGT TTTTGTGAGT TTTTGTGAGT TTTTGTGAGT  
 3081 CTGCTGCTTG CAAACAAA AACCACCGCT ACCACCGCT ACCACCGCT ACCACCGCT ACCACCGCT ACCACCGCT ACCACCGCT  
 3151 CTTTTCCTGA AGGTAACTGG CTTACAGCAG CCGCAGATAC CCGCAGATAC CCGCAGATAC CCGCAGATAC CCGCAGATAC  
 3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CCGCTACATA CCGCTACATA CCGCTACATA CCGCTACATA CCGCTACATA  
 3291 TGCTGCCAGT GCGGATAAGT CGTGTCTTAC CCGGTTGGAC CCGGTTGGAC CCGGTTGGAC CCGGTTGGAC CCGGTTGGAC  
 3361 CCGTCCGGCT GAACGGGGG TTTGTGACCA CAGCCAGCT TGGAGCGAAC TGGAGCGAAC TGGAGCGAAC TGGAGCGAAC  
 3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CCGTTCCTGA AGGAGAAAG AGGAGAAAG AGGAGAAAG AGGAGAAAG  
 3501 CCGCAGGGTC GGAACAGGAG AGCGCACGAG GAGCTTCTCA GAGCTTCTCA GAGCTTCTCA GAGCTTCTCA GAGCTTCTCA  
 3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTNGT GATGCTCGTC AGGGGGCGG CCTGGTATCT TTATAGTCTT  
 3641 AAAACGCCAG CAACGCGGCT TTTTACCGT TTTTACCGT TTTTACCGT TTTTACCGT TTTTACCGT TTTTACCGT  
 3711 TGGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC  
 3781 CGAACGACCG AGCGACGGA GTCAGTGAAG GAGGAGCGG AAGAGCGCC CCGGTTATCC CCGGTTATCC CCGGTTATCC  
 3851 CCGCGCGTTG GCGGATTCAT TAAATGACGT GGCACGACAG GTTTCCTGAC TGGAAAGCGG TGGAAAGCGG GCAGTGAGCG

FIG.-32D

```

3921 CAACGCAATT AATGTGAGTT AGCTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC

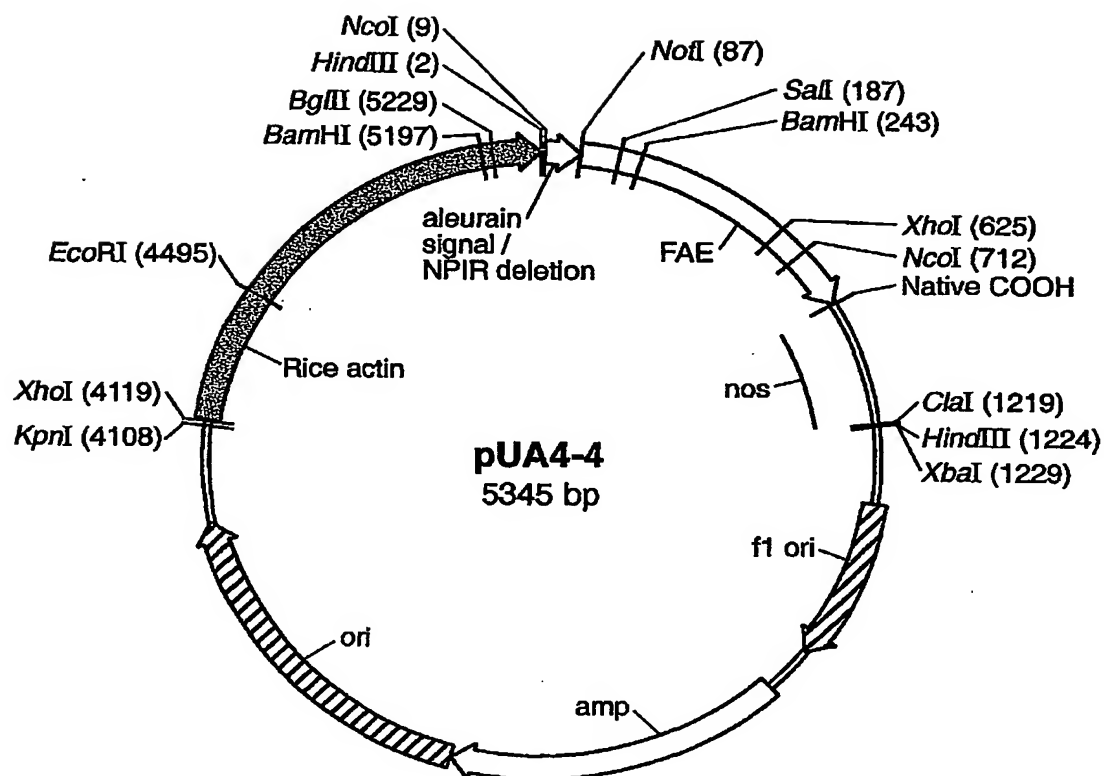
                               XhoI
4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCCC CTCGAGGTCA TTCATATGCT
4131 TGAGAAGAGA GTCGGGATAG TCCAAAATAA AACAAAGGTA AGATTACCTG GTCAAAAGTG AAAACATCAG
4201 TTAAAAGGTT GTATAAGTAA AATATCGGTA ATAAAGGTTG GCCCAAAGTG AAATTACTC TTTTCTACTA
4271 TTATAAAAAT TGAGGATGTT TTGTCGGTAC TTTGATACGT CATTTTGTGA TGAATTGGTT TTTAAGTTTA
4341 TTCGCGATTT GGAATGCAAT ATCTGTATTT GAGTCGGTTT TTAAGTTTCGT TGCTTTTGTG AATACAGAGG
4411 GATTTGTATA AGAATATATCT TTAATAAACCC CATATGCTAA TTTGACATAA TTTTGTGAAA AAATATATAT

                               EcoRI
4481 TCAGGCGGAAT TCCACAAATGA ACAATAATAA GATTAAATA GCITGGCCCC GCITGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAGATA AACTTAGACT CAAAACATTT ACAMAAACAA CCCCTAAAGT CCTAAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCAACCC AACCCACCCC AGTGCAGCCA
4691 ACTGGCAAAAT AGTCTCCACC CCCGGCACTA TCACCGTGAG TTGTCCGAC CACCGCACGT CTCGCAGCCA
4761 AAAAAAATAA AAGAAAGAAA AAAAAAATAA AGAAAACAG CAGGTGGGTC CGGTCGTGG GGGCCGAAA
4831 AGCGAGGAGG ATCGCGAGCA GCGACGAGGC CCGGCCCTCC CTCCGCTTCC AAAGAAACGC CCCCCTCGC
4901 CACTATATAC ATACCCCCC CTCTCTCTCC ATCCCCCAA CCTACCAAC ACCACCAACA CCACCTCTC
4971 CCCCCTGGCT GCCGGACGAC GAGCTCCTCC CCCCTCCCC TCAGCCGCGG CCGGTAACCA CCCCCTCCCT
5041 CTCTCTCTTC TTCTCTCGTT TTTTCTCTCG TCTCGGTCCT GATCTTTGGC CTGTGTAGTT TGGGTGGGCG
5111 AGAGCGGCTT CGTCGCCAG ATCGGTGCGC GGGAGGGGCG GATCTCGCG GCTGGCGTCT CCGGGCGTGA

                               BamHI
5181 GTCGGCCCCG ATCTCTCGCG GGAATGGGGC TCTCGGATGT AGATCTCTT TCTTTCTTCT TTTTGTGTA
5251 GAATTTGAA CTCTCAGCAT TGTTCATCGG TAGTTTTTCT TTTTCATGAT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTGTAGC

```

FIG. 32E

**FIG.\_33A**

```

NGOI
~~~~~
HindIII
~~~~~
      M A E A R V L L L A L A V L A T A A V A V
1  AAGCTTACCA TGGCCACAGC CCGGTCCTC CTCTGGCGC TCGCCGTGCT GGCACGGCC GCCGTGCGCG

      ~~~~~
 . A S S R A A A S T Q G I S E D L Y S R L V E M .
71 TCGCCTCCTC CCGGCGGCC GCCTCCAGC AGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAAT

      ~~~~~
      . A T I S Q A A A Y A D L C N I P S T I I K G E K
141 GGCCACTATC TCCCAAGCTG CCTACGCCGA CTGTGCAAC ATTCCGTGCA CTATTATCAA GGGAGAGAAA

      ~~~~~
 I Y N S Q T D I N G W I L R D D S S K E I I T V
211 ATTTACAATT CTCAACTGA CATTACGGA TGGATCCTCC GCGACGACAG CAGCAAGAA ATAATCACCG

 . F R G T G S D T N L Q L D T N Y T L T P F D T .
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAACTAC ACCCTCAGC CTTTCGACAC

 . L P Q C N G C E V H G Y Y I G W V S V Q D Q
351 CCTACCACAA TGCAACGGTT GTGAAGTACA CCGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA

 V E S L V K Q Q V S Q Y P D Y A L T V T G H X L
421 GTCGAGTCGC TTGTCAAACA GCAGGTAGC CAGTATCCGG ACTACGGCT GACCGTGACC GGCCACKCCC

 . G A S L A A L T A A Q L S A T Y D N I R L Y T .
491 TCGGCGCCTC CCTGGCGGCA CTCACTGCCG CCCAGCTGTC TCGGACATAC GACAACATCC GCCTGTACAC

      ~~~~~
      . F G E P R S G N Q A F A S Y M N D A F Q A S S
561 CTTGGCGGAA CCGCGCAGCG GCAATCAGCG CTTGCGGTCC TACATGAACG ATGCTTTCCA AGCCTCGAGC

      P D T T Q Y F R V T H A N D G I P N L P F V E Q
631 CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAAGC ACGGCATCCC AAACCTGCCC CCGGTGGAGC

      ~~~~~
 . G Y A H G G V E Y W S V D P Y S A Q N T F V C .

```

FIG.\_33B



|      |            |             |             |            |             |             |            |
|------|------------|-------------|-------------|------------|-------------|-------------|------------|
| 701  | AGGGGTACGC | CCATGGCGGT  | GTAGAGTACT  | GGAGCGTTGA | TCCTTACAGC  | GCCCAGAAC   | CATTGTCTG  |
|      | T G D      | E V Q C     | E A         | Q G G      | Q G V N     | N A H       | T T Y      |
| 771  | CACGTGGGAT | GAAGTGCAGT  | GCTGTGAGC   | CCAGGGCGGA | CAGGGTGTGA  | ATAATGCGCA  | CACGACTTAT |
|      | F G M T    | S G A       | C T W       | *          |             |             |            |
| 841  | TTTGGGATGA | CGAGCGGAGC  | CTGTACATGG  | TGATCAGTCA | TTTCAGCCCTC | CCCGAGTCTA  | CCAGGAAAGA |
| 911  | TGGATGTCTT | GGAGAGGGGG  | CCGCTTAACC  | ACTGARGGAT | GAGCTGTAAA  | GAAGCAGATC  | GTTCAAACAT |
| 981  | TTGGCAATAA | AGTTTCTTAA  | GATTGAATCC  | TGTTGCCGGT | CTTGCCGATGA | TTATCATATA  | ATTTCTGTGG |
| 1051 | AATTACGTTA | AGCATGTAAAT | AATTAACATG  | TAATGCTATG | CGTTATTTAT  | GAGATGGGTT  | TTTATGATTA |
| 1121 | GAGTCCCGCA | ATTATACAT   | TAATACGCGA  | TAGAAACAA  | AATATAGCGC  | GCAAACTAGG  | ATAAATTATC |
|      |            |             |             | HindIII    |             |             |            |
|      |            |             | Clal        | XbaI       |             |             |            |
| 1191 | GCGCGCGGTG | TCATCTATGT  | TACTAGATCG  | ATAAGCTTCT | AGAGCGGCCG  | GTGGAGCTCC  | AATTCGCCCT |
| 1261 | ATAGTGAGTC | GTATTACGG   | CGCTCACTGG  | CCGTCTGTTT | ACAACGTCTG  | GACTGGGAAA  | ACCTTGGCGT |
| 1331 | TACCCAACTT | AATCGCTTG   | CAGCACATCC  | CCCTTTCGCC | AGCTGGCGTA  | ATAGCGAAGA  | GGCCCGCACC |
| 1401 | GATCGCCCTT | CCCAACAGTT  | GCGCAGCCTG  | AATGGCGAAT | GGGACGGGCC  | CTGTAGCGGC  | GCATTAAAGC |
| 1471 | CGCGGGGTGT | GGTGGTTACG  | CGCAGCGTGA  | CCGCTACACT | TGCCAGCGCC  | CTAGCGCCCG  | CTCCTTTTCG |
| 1541 | TTTCTTCCCT | TCCTTCTCG   | CCACGTTTCG  | CGGCTTCC   | CGTCAAGCTC  | TAAATCGGGG  | GCTCCCTTTA |
| 1611 | GGGTTCGGAT | TTAGTGTCTT  | ACGGCACCTC  | GACCCCAAA  | AACTTGATTA  | GGGTGATGGT  | TCACGTAGTG |
| 1681 | GGCCATCGCC | CTGATAGACG  | GTTTTTCGCC  | CTTTGACGTT | GGAGTCCACG  | TTCTTTAATA  | GTGGACTCTT |
| 1751 | GTTCCTAACT | GGAAACAAC   | TGAGCTGATT  | TAACAAAAAT | TTAACGCGAA  | TATAAGGGAT  | TTTGGCCGAT |
| 1821 | TCGGCCTATT | GGTTAAAAA   | TTGAGCTGAT  | TAACAAAAAT | TTAACGCGAA  | TTTAAACAAA  | ATATTAACGC |
| 1891 | TTCAATATTA | GGTGGCATT   | TTGCGGGGAA  | TGTGCGCGGA | ACCCCTATTT  | GTATTATTTT  | CTAAATACAT |
| 1961 | TCAATATATG | ATCCGCTCAT  | GAGACAATAA  | CCCTGATATA | TGCTTCAATA  | ATATTGAAAA  | AGGAAGAGTA |
| 2031 | TGAGTATTCA | ACATTTCCGT  | GTCCGCCCTTA | TTCCCTTTT  | TGCGGCAATTT | TGCTTCTCTG  | TTTGTGCTCA |
| 2101 | CCAGAAACG  | CTGGTGAAG   | TAAAGATGC   | TGAAGATCAG | TTGCGGTCAC  | GAGTGGGTTA  | CATCGAATG  |
| 2171 | GATCTCAAC  | CGGTAAGAT   | CCTTGAGAGT  | TTTCGCCCCG | AAGAAGCTTT  | TCCAATGATG  | AGCACITTTA |
| 2241 | AAGTCTGCT  | ATGTGGCGCG  | GTATTATCCC  | GTATTGACGC | CGGGCAAGAG  | CAACTCGGTC  | GCCGCATACA |
| 2311 | CTATTCTCAG | AATGACTTGG  | TTGAGTACTC  | ACCAGTCACA | GAAAGCATC   | TTACGGATGG  | CATGACAGTA |
| 2381 | AGAGAATTAT | GCAGTGTGTC  | CATAACCATG  | AGTGATAACA | CTCGGCCCAA  | CTTACTTCTG  | ACAAAGATCG |
| 2451 | GAGAACCGAA | GGAGCTAACC  | GCCTTTTTCG  | ACAACATGGG | GGATCATGTA  | ACTCGCCTTG  | ATCGTTGGGA |
| 2521 | ACCGGAGCTG | AATGAAGCCA  | TACCAAAACGA | CGAGCGTGAC | ACCACGATGC  | CTGTAGCAAT  | GGCAACAACG |
| 2591 | TTGCGCAAAC | TATTAACTGG  | CGAACTACTT  | ACTCTAGCTT | CCGGCAACA   | ATTAATAGAC  | TGGATGGAGG |
| 2661 | CGGATAAAGT | TGCAGGACCA  | CTTCTGCGCT  | CGGCCCTTCC | GGCTGGCTGG  | TTTATTTGCTG | ATAAATCTGG |

FIG. 33C

2731 AGCCGGGTGAG CGTGGGTCTC GCGGTATCAT TGCAGCAGTG GGGCCAGATG GTAAGCCCTC CCGTATCGTA  
 2801 GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCCT  
 2871 CACTGATTAA GCATTGGTAA CTGTGAGACC AAGTTTACTC ATATATACTT TAGATTGATT TAAAACTTCA  
 2941 TTTTAAATTT AAAAGGATCT AGGTGAAGAT CTTTTTGAT AATCTCATGA CCAAAATCCC TTAACGTGAG  
 3011 TTTTCGTTC ACTGAGCTC AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT TTTTCTGC  
 3081 GCCTAATCTG CTGCTTGCAA ACAAATAAAC CACCGCTACC AGCGGTGGTT TGTTCGCGG ATCAAGAGCT  
 3151 ACCAACTCTT TTCCGAGAG TAAC'TGGCTT CAGCAGAGCG CAGATACCAA ATACTGTCTT TCTAGTGTAG  
 3221 CCGTAGTTAG GCCACCACTT CAAGAATCTT GTAGCACCGG CTACATACCT CGCTGTGCTA ATCCTGTTAC  
 3291 CAGTGGCTGC TGCACGTGGC GATAAGTCTT GTCTTACCGG GTTGGACTCA AGACGATAGT TACC GGATAA  
 3361 GGCGCAGCGG TCGGCTGAA CGGGGGGTTC GTGCACACAG CCCAGCTTGG AGGAAAGCGG CTACACCGAA  
 3431 CTGAGATACC TACAGCGTGA GCTATGAGAA AGCGCCACGC TTCCCGAAGG GAGAAAGCGG GACAGGTATC  
 3501 CGGTAAGCGG CAGGCTCGGA ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGTTCGTGAT GGTATCTTTA  
 3571 TAGTCTGTC ACGCCAGCAA CGCGGCTTTT TTAGCGTTCC TGGCTTTTG GCTCACATGT GGTTCGTGAT  
 3641 CTATGGAAA ACCTTCGCC ACCTTCGACT TGAGCGTCCA TTTTGTGAT GCTGCTTTT CTGACATGT  
 3711 TCTTCTGTC GTTATCCCTT GATTCGTGG ATAAACGTAT TACCAGCTTT GAGTGAGCTG ATACCGCTCG  
 3781 CCGCAGCCGA ACGACTGAGC GCAGCGAGTC AGTGAGCGAG GAAGCGGAAG AGCGCCCAT ACGCAAACCG  
 3851 CCTCTCCCG CGGTTGGCC GATTCAATAA TGACAGTGGC ACGACAGGTT TCCCGACTGG AAGCGGGCA  
 3921 GTGAGCGCAA CGCAATTAAT GTGAGTTAG TCACTCATTA GGCACCCAG GCTTTACACT TTATGCTTCC  
 3991 GGCTCGTATG TTGTGTGAA TTGTGAGCGG ATAAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC  
  
 4061 GCCAAGCGCG CAATTAACCC TCACTAAAGG GAACAAAAGC TGGGTACCGG GCGCCCTCTC GAGGTCAATC  
 4131 ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAAC AAAGGTAAGA TTACCTGGTC AAAAGTGAAA  
 4201 ACATCAGTTA AAAGGTGGTA TAAGTAAAT ATCGGTAATA AAAGGTGGCC CAAAGTGAAA TTACTCTTTT  
 4271 TCTACTATTA TAAAAATTGA GGATGTTTGG TCGGTACTTT GATACGTCTT TTTTGTATGA ATTGGTTT  
 4341 AAGTTTATTC GCGAATTGGA AATGCATATC TGTATTTAG TCGGTTTAA AGTTCGTTGC TTTTGTAAAT  
 4411 ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAACCCCAT ATGCTAATTT GACATAAATTT TTGAGAAAAA  
  
 4481 TATATATTCA GCGAATTTCC ACAAATGACA ATAATAAGAT TAAAAAGCT TGCCCCCGTT GCAGCGATGG  
 4551 GTATTTTTC TAGTAAATA AAAGATAAAC TTAGACTCAA AACATTACA AAAACAACCC CTAAAGTCTT  
 4621 AAAGCCCAAA GTGCTATGCA CGATCCATAG CAAGCCCAAG CCAACCCCAAC CCAACCCCAAGT  
 4691 GCAGCCCAACT GGCAATAGT CTCCACCCCG GGCACATCA CCGTGAGTTG TCCGCAACCC CCGACGCTC  
 4761 GCAGCCCAAA AAAAAAAG AAGAAAAAAG AAGAAAAAAG AAGAAAAAAG GTGGGTCCGG GTCGTGGGG  
 4831 CCGAAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCGCG GCGCTCCCTC CGCTTCCAAA GAAACGCCCC

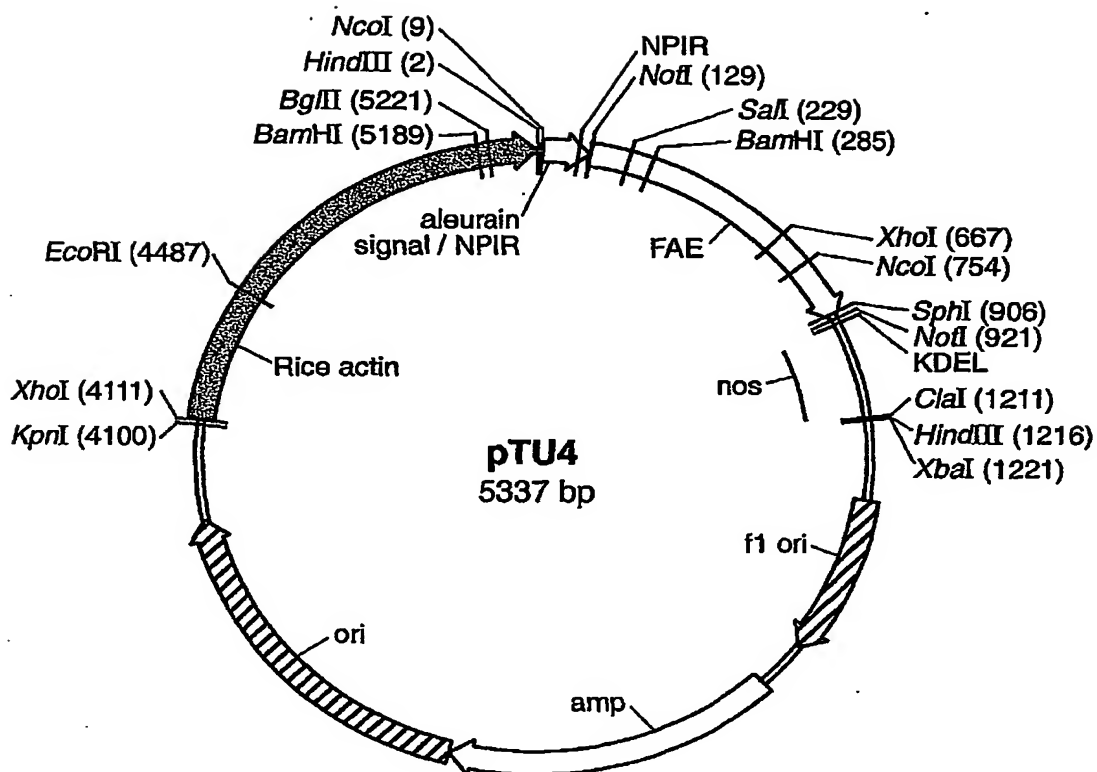
FIG.-33D

```

4901 CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCACCC TACCACCAAC ACCACCAACA
4971 CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCTCCCCC CTCCTCTCC GCCGCCGCC GTAACCAACC
5041 CGCCCCCTCTC CTCCTTCTTT CTCCTTTTTT TTTTTCGTC CCGTCTCGAT CTTTGGCCCTT GGTAGTTTGG
5111 GTGGGCGAGA GCGGCTTCGT CCCCCAGATC GGTGCGCGGG AGGGCGGGA TCTCGCGGT GCGTCTTCCG
 BamHI
5181 GCGGTGAGTC GGCCCGGATC CTCGCGGGA ATGGGCTCT CGATGTAGA TCTTCTTCTT TTCTTCTTTT
5251 TGTGGTAGAA TTGAATCCC TCAGCATTGT TCATCGGTAG TTTTCTTTT CATGATTTGT GACAAATGCA
5321 GCGTCTGCG GAGCTTTTT GTAGC
 BglII

```

**FIG.\_33E**

**FIG. 34A**

NcoI  
 HindIII  
 M A H A R V L L L A L A V L A T A A V A V  
 1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGCG TCGCGTGT GGCACGGCC GCGTGGCGG  
 NotI  
 . A S S S F A D S N P I R P V T D R A A A S T .  
 71 TCGCTCCTC CTCTCCTTC GCCGACTCCA ACCGATCG GCCGTCACC GACCGCGCG CCGCTCCAC  
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A  
 141 GCAGGGCATC TCCGAAGACC TCTACAGCG TTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC  
 SalI  
 D L C N I P S T I I K G E K I Y N S Q T D I N G  
 211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTAACG  
 BamHI  
 . W I L R D D S S K E I I T V F R G T G S D T N .  
 281 GATGGATCCT CCGGACGAC AGCAGCAAG AATAATCAC CGTCTTCGT GGCACGTGTA GTGATACGAA  
 . L Q L D T N Y T L T P F D T L P Q C N G C E V  
 351 TCTACAATC GATACTAAT ACACCTCACC GCCTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA  
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
 421 CACGGTGGAT ATTATATTGG ATGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA CAGCAGGTTA  
 . Q Y P D Y A L T V T G H X L G A S L A L T A .  
 491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG CACTCACTGC  
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q  
 561 CGCCCAAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG  
 XhoI  
 A F A S Y M N D A F Q A S S P D T T Q Y F R V T  
 631 GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCTCGA GCCCAGATAC GACGCGAGTAT TTCCGGGTCA  
 NcoI  
 . H A N D G I P N L P P V E Q G Y A H G G V E Y .

FIG. 34B

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCAGTGGG GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 . W S V D P Y S A Q N T F V C T G D E V Q C C E  
 771 CTGGAGCGTT GATCCTTACA GCGCCAGAA CACATTGTC TGCACGGGG ATGAAGTGA GTGCTGTGAG  
 SphI  
 A Q G G Q G V N N A H T T Y F G M T S G A C T W  
 841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT  
 NotI  
 . P V A A A E P L K D E L \*  
 911 GGCGGTGCG GCGCGCGGAA CCACTGAAGG ATGAGCTGTA AAGAAGCAGA TCGTTCAAC ATTTGGCAAT  
 981 AAAGTTCTT AAGATTGAAT CCTGTTGCCG GTCTTGCGAT GATATCATTA TAATTTCTGT TGAATTACGT  
 1051 TAAGCATGTA ATAATTAA CA TGAATGCAAT GACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG  
 1121 CAATTATACA TTTAATACGC GATAGAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG  
 HindIII  
 ClaI XbaI  
 1191 TGTCACTCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC CTATAGTGAG  
 1261 TCGTATFACG CCGCTCACT GCGCGTCTGT TTACAACGTC GTGACTGGGA AAACCTCGGC GTTACCCAAAC  
 1331 TTAATCGCCT TGCAGCACAT CCCCCTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC  
 1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG CGGCATTAG CGCGCGCGGT  
 1471 GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTTGCCAGCG CCTAGCGCC CGCTCTTTC GCTTCTTTC  
 1541 CTTCCCTTCT TTACGGCACC TCGACCCCAA AAACTTGT TAGGTGATG GGGCTCCCTT TAGGGTTCCG  
 1611 ATTTAGTGCT TTACGGCACC TCGACCCCAA AAACTTGT TAGGTGATG GGGCTCCCTT TAGGGTTCCG  
 1681 CCCTGATAGA CGGTTTTCG CCCTTTCG TCGACCCCAA AAACTTGT TAGGTGATG GGGCTCCCTT TAGGGTTCCG  
 1751 CTGGAAACAAC ACTCAACCTT ATCTCGGTCT ATCTCTTTCG TTTAAGGGG ATTTTGGCGA TTTTCGGCCTA  
 1821 TTGGTTAAAA AATGAGCTGA TTTAACAATA ATTTAAGCG AATTTAACA AATATAAAT GCTTACAATT  
 1891 TAGGTGGCAC TTTTCGGGGA AATGTCGCG GAACCCCTAT TGTATTATTT TTCTAAATAC ATTCAAATAT  
 1961 GTATCCGCTC GTGTGCCCCT TATTCCTTTC AATGCTTCAA TAATATTGAA AAAGGAAGAG TATGAGTATT  
 2031 CAACATTTCC GTGTGCCCCT TATTCCTTTC TTTTCGGGCT TTTGCTTTC TGTTTTGT CACCCAGAAA  
 2101 CGCTGGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAAC TGGATCTCAA  
 2171 CAGCGGTAAG ATCTTTGAGA GTTTTCGCC CCAAGAACGT TTTCCTTTC TGAAGTCTG TGAAGTCTG  
 2241 CATGTGGCG CGGTATTATC CCGTATTGAC GCGGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC  
 2311 AGAATGACTT GGTGTAGTAC TCACCACTAC CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGAGATT  
 2381 ATGAGTGTCT GCCATAACCA TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG

FIG.-34C

2451 AAGGAGCTAA CCGCTTTTTC GCACAACATG GGGGATCATG TAACTCGCCT TGATCGTTGG GAACCGGAGC  
 2521 TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA  
 2591 ACTATTAACT GCGGAACTAC TTAATCTAGC TTCCCGGCAA CAATTAAATG ACTGGATGGA GCGCGATAAA  
 2661 GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT CCGTTATTCG TGATAAATCT GGAGCCGGTG  
 2731 AGCGTGGGTC TCBCGATATC ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCGATATCG TAGTTATCTA  
 2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG AGATAGGTGC CTCACATGAT  
 2871 AAGCATTTGT AACTGTGAGA CCAAGTTTAC TCATATATAC TTTAGATTGA TTTAAAACCTT CATTTTAAAT  
 2941 TAAAAGGAT CTAGGTGAAG ATCTTTTGTG ATAATCTCAT GACCAAAATC CCTTAAACGTG AGTTTTCGTT  
 3011 CCATGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT CATTGAGATC TTTTTCGCTT GCGCGTAAATC  
 3081 TGCTGCTTGC AAACAATAAA GGTAACTGGC TTCAGCAGAG CCAGATATAC CTTGCTGTGC TAACTCTGTT ACCAGTGGCT  
 3151 TTTTTCGAA AGGCCACCCAC GTTAACTGGC TTCAGCAGAG CCAGATATAC CTTGCTGTGC TAACTCTGTT ACCAGTGGCT  
 3221 AGGCCACCCAC GTTAACTGGC TTCAGCAGAG CCAGATATAC CTTGCTGTGC TAACTCTGTT ACCAGTGGCT  
 3291 GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGTGTGACT GGTGTGACT GGTGTGACT GGTGTGACT GGTGTGACT  
 3361 GGTGCGGCTG AACGGGGGT TCGTGACAC AGCCAGCTT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT  
 3431 CCTACAGCT GAGCTATGAG AAAGCGCCAC GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT  
 3501 GGCAGGCTCG GAACAGGAGA GGCAGGAG GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT  
 3571 TCGGCTTTCG CCACCTCTGA CTTGAGCGTC GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT  
 3641 AACGCCAGC AACGCGGCTT TTTTACGGTT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT  
 3711 GCGTATCCG CCGATCTCTG GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT GGTGACT  
 3781 GAACGACCGA GCGCAGCGAG TCACTGAGCG AGGAAGCGG AGGAAGCGG AGGAAGCGG AGGAAGCGG AGGAAGCGG  
 3851 CCGCGGTTGG CCGATCTCAT AATGCACTG GCGCAGCGAG AGGAAGCGG AGGAAGCGG AGGAAGCGG AGGAAGCGG  
 3921 AACGCAATTA ATGTGAGTTA GCTCACTCAT TAGGCACCCC AGGCTTTACA CTTTATGCTT CCGGCTCGTA  
 3991 TGTGTGTGG AATGTGAGC GGTAAACAAT TTCACACAGG AACAGCTAT GACCATGATT ACGCCAAGCG

KpnI XhoI

4061 CGCAATTAACT CCTCACTAAA GGGAAACAAA GCTGGGTACC GGGCCCCCCC TCGAGGTGAT TCATATGCTT  
 4131 GAGAAAGAG TCGGGAATAGT CCAAAATAAA ACAAGGTAA GATTACCTGG TCAAAAGTGA AAACATCAGT  
 4201 TAAAAGGTGG TATAAGTAAA ATATCGGTAA TAAAAGGTGG CCAAAAGTGA AATTTACTCT TTTCTACTAT  
 4271 TATAAAATTT GAGGATGTTT TGTCTGTTACT TTTGATACGC ATTTTGTGAT GAATTGGTTT TTAAGTTTAT  
 4341 TCGCGATTGG GAAATGCATA TCTGTATTGG AGTCGGTTT TAAATTCGTT TTTTGTGAAA ATACAGAGGG  
 4411 ATTGTATATA GAAATATCTT TAAAACCC ATATGCTAAT TTGACATAAT TTTTGTGAAA AATATATATT

EcoRI

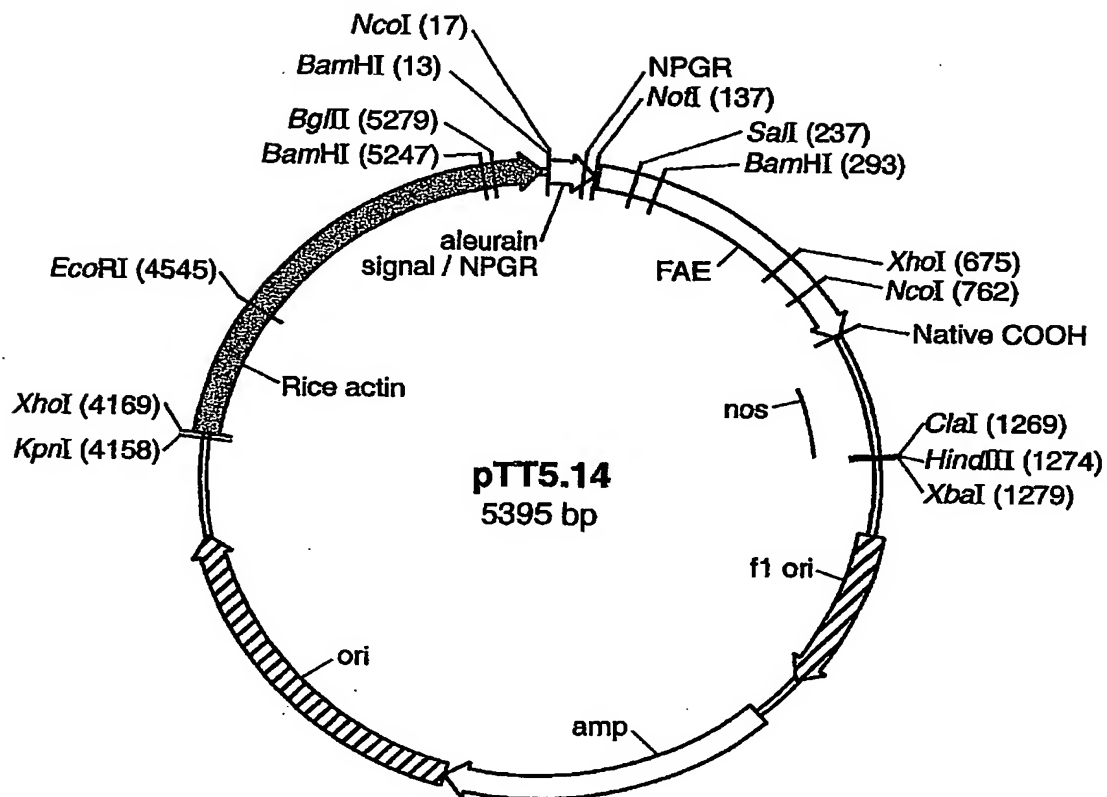
4481 CAGGCGAATT CCACAATGAA CAATAATAAG ATTAATAATAG CTTGCCCCCG TTGCAGCGAT GGGTATTTT  
 4551 TCTAGTAAAA TAAAAGATAA ACTTAGACTC AAACATTTA CAAAACAC CCCTAAAGTC CTAAGGCCCA

FIG.-34D

4621 AAGTGTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCACCCCA GTGCAGCCAA  
4691 CTGGCAATA GTCTCCACCC CCGGCACATAT CACCGTGAGT TGTCCGCACC ACCGCACGTC TCGCAGCCAA  
4761 AAAAAAAAAA AGAAGAAAA AAAAAAAAAA CACCGTGAGT AGGTGGGTCC AGGTCTGTGGG GGCCGGAAAA  
4831 GCGAGGAGGA TCGCGAGCAG CGACGAGGCC CGGCCCTCCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC  
4901 ACTATATACA TACCCCCCCC TCTCCTCCCA TCCCCCCTAC CTAACACCA CCACCAACCAC CACCTCCTCC  
4971 CCCCCTGCTG CCGGACGACG AGCTCCTCCC CCTCCCTCC CCGCCGCCGC CGGTAAACCAC CCGCCCTCTC  
5041 TCCCTCTTCT TTCTCCGTTT TTTTCTTCTG CTGGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCGA  
5111 GAGCGGCTTC GTGCCCCAGA TCGETCCCG GAGGGGCGG GATCTCGCGG CTGGCTCTC CGGGCGTGAG  
BamHI BglII  
\*\*\*\*\*  
5181 TCGGCCCGGA TCCTCGCGG GAATGGGGCT CTCGATGTA GATCTCTTT CTTCCTCTT TTGTGGTAG  
5251 AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTTCTT TTCATGATTT GTGACAAATG CAGCCTCGTG  
5321 CGGAGCTTTT TTGTAGC

FIG.\_34E



**FIG.\_35A**

```

 NgoI
      ~~~~~
      BamHI
      ~~~~~
1 CCTGACGCG AGGATCCATG GCCACGCCC GCGTCTCTCT CCTGGGCTC GCCGTGCTGG CCACGGCGCG
 . V A V A S S S F A D S N P G R P V T D R A A
71 CGTCGGCGTC GCCTCCTCTT CCTCCTTCG CCACTCCAAC CCGGGCCGGC CCGTCACCGA CCGCGCGCGC
 NotI
      ~~~~~
141  A S T Q G I S E D L Y S R L V E M A T I S Q A A
      Sali
      ~~~~~
211 . Y A D L C N I P S T I I K G E K I Y N S Q T D .
 BamHI
      ~~~~~
281  . I N G W I L R D D S S K E I I T V F R G T G S
      D T N L Q L D T N Y T L T P F D T L P Q C N G C
351  GATACGAATC TACAACTCGA TACTAACTAC ACCCTCAGC CTTTCGACAC CCTACCACAA TGCAACGGTT
      . E V H G G Y Y I G W V S V Q D Q V E S L V K Q .
421  GTGAAGTACA CCGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA GTCGAGTCGC TTGTCAAACA
      . Q V S Q Y P D Y A L T V T G H X L G A S L A A
491  GCAGGTTAGC CAGTATCCGG ACTACGGGCT GACCGTGACC GGCACKCCC TCGSCGCCCTC CCTGGCGGCA
      L T A A Q L S A T Y D N I R L Y T F G E P R S G
561  CTCACGTGCG CCCAGCTGTC TCGGACATAC GACAACATCC GCCTGTACAC CTTTCGGCGAA CCGCGCAGCG
      XhoI
      ~~~~~
631 . N Q A F A S Y M N D A F Q A S S P D T T Q Y F .
 GCAATCAGGC CTTTCGGTCTG TACATGAACG ATGCCTTCCA AGCCTCGAGC CCAGATACGA CGCAGTATTT
 NgoI
      ~~~~~
      . R V T H A N D G I P N L P P V E Q G Y A H G G

```

FIG.\_35B

701 CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCC CCGGTGGAGC AGGGGTACGC CCATGGCGGT  
 V E Y W S V D P Y S A Q N T F V C T G D E V Q C  
 771 GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACCA CATTTGTCTG CACTGGGGAT GAAGTGCAGT  
 . C E A Q G G Q G V N A H T T Y F G M T S G A .  
 841 GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATATGCGCA CACGACTTAT TTTGGGATGA CGAGCGGAGC  
 . C T W \*  
 911 CTGTACATGG TGATCAGTCA TTTCAGCCTC CCGAGTGTA CCAGGAAGA TGGATGTCTT GGAGAGGGGG  
 981 CCGGTAACC ACTGAAGGAT GAGCTGTAA GAAGCAGATC GTTCAACAT TTGGCAATAA AGTTTCTTAA  
 1051 GATTGAATCC TGTGCGCGGT CTTCGATGA TTATCATATA ATTCTGTG AATTACGTTA AGCATGTAAT  
 1121 AATTACATG TAATGCATGA CGTTATTAT GAGATGGGT TTTATGATTA GAGTCCCGCA ATTATACATT  
 1191 TAATACCGA TAGAAACAA AATATAGCGC GCAACTAGG ATAAATTATC GCGCGCGGTG TCATCTATGT  
 HindIII  
 ~~~~~  
 ClaI XbaI
 ~~~~~  
 1261 TACTAGATCG ATAAGCTTCT AGAGCGGCGG GTGGAGCTCC AATTCGCCCT ATAGTGAGTC GTATTACGGC  
 1331 CGCTCACATGG CCGTCGTTT ACACCGTCTG GACTGGGAA ACCCTGGCGT TACCCAACCTT AATCGCCTTG  
 1401 CAGCACATCC CCTTTCGCC AGCTGGCGTA CTGCGAGA GGGCCGCACC GATCGCCCTT CCCAACAGTT  
 1471 GCGCAGCCTG AATGGCGAAT GGGACGCGCC ATAGTAGCGC GCATTAAGCG CCGCGGGTGT GGTGGTTACG  
 1541 CGCAGCGTGA CCGCTACACT TGGCAGCGCC CTAGCGCCCG CTCTTTTCGC TTTCTTCCCT TCCTTTCTCG  
 1611 CCACGTTGCG CGGCTTTCC CGTCAAGCTC TAAATCGGG GCTCCCTTTA GGGTTCCGAT TTAGTGCTTT  
 1681 ACGGCACCTC GACCCCAAA AACTTGATTA GGGTGATGGT TCACGTAGTG GGGCATCGCC CTGATAGACG  
 1751 GTTTTTCGCC CTTTGACGTT GGAGTCCACG TTCTTTAATA GTGGACTCTT GTTCCAACT GGAACAACAC  
 1821 TCAACCCCTAT CTCGGTCTAT TCTTTTGATT TATAAGGGAT TTGCGCGATT TCGGCCCTATT GGTAAAAA  
 1891 TGAGCTGATT TAACAAAAAT TTAACGCGAA TTTTAAACAA ATATTAACGC TTACAATTTA GGTGGCACCT  
 1961 TTCGGGGAAA TGTGCGCGGA ACCCTATT GTTTATTTT CTAAATACAT TCAATATATG ATCCGCTCAT  
 2031 GAGACAATAA CCTGATATA TGTTCATAA ATATTGAAA AGGAAGAGTA TGAATATTCA ACATTTCCGT  
 2101 GTCGCCCTTA TTCCCTTTT TCGGGCATTT TGCCTTCCCT CATCTCAACA CCCAGAAACG CTGGTGAAG  
 2171 TAAAGATGC TGAAGATCAG TTGGGTGCAC GAGTGGGTTA GATCTCAACA GATCTCTGCT ATGTGGCGCG  
 2241 CCTTGAGAGT TTTCGCCCG AAGAAGCTT TCCAATGATG AGCATTITA AGTTCTCTAG AATGACTTGG  
 2311 GTATTATCCC GTATTGACGC CCGGCAAGAG CAATCGGTC CCGCATACA CTATTTCTAG AATGACTTGG  
 2381 TTGAGTACTC ACCAGTCACA GAAAGCATC TTACGGATGG CATGACAGTA AGAGATTAT GCAGTGTGCG  
 2451 CATAAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG ACAACGATCG GAGGACCGAA GGAGCTAACCC  
 2521 GCTTTTITGC ACAACATGGG GGATCATGTA ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA  
 2591 TACCAAACGA CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAACG TTGCGCAAC TATTAACTGG  
 2661 CGAACTACTT ACTCTAGCTT CCCGGCAACA ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA

FIG.-35C

2731 CTTCTGGCGT CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG CGTGGGTCTC  
 2801 GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG  
 2871 TCAGGCAACT ATGGATGAAC GAAATAGACA GATCGCTGAG ATAGGTGCTT CACTGATTAA GCATTGGTAA  
 2941 CTGTGAGACC AAGTTTACTC ATATATACTT TAGATTGATT TAAAACTTCA TTTTAAATTT AAAAGGATCT  
 3011 AGGTGAAGAT CCTTTTGGAT AATCTCATGA CCAAAATCCC TTAACGTGAG TTTTCGTTC ACTGAGCGTC  
 3081 AGACCCCGTA GAAAAGATCA AAGGATCTTC TTGAGATCCT TTTTCTGTC GCGTAATCTG CTGCTTGCAA  
 3151 ACAAAAACAC CACCGTACC AGCGGTGGTT TGTTGGCCGG ATCAAGAGCT ACCAACTCTT TTTCCGAAGG  
 3221 TAATGGCTT CAGCAGAGCG CAGATACCA ATACTGCTT TCTAGTGTAG CCCACCACTT  
 3291 CAAGAACTCT GTAGCACCGC CTACATACCT GCTCTGTCTA ATCCTGTTAC CAGTGGCTGC TGCCAGTGGC  
 3361 GATAAGTCGT GTCTTACCGG GTTGGACTCA AGACGATAGT TACCGGATAA CCGCGAGCGG TCGGGGTGAA  
 3431 CCGGGGGTTC GTGCACACAG CCCAGCTTGG AGCGAACGAC CTACACCGAA CTACAGCTGA  
 3501 GCTATGAGAA AGCGCCACGC TTCCCGAAG GAGAAAGCG GACAGGTATC CCGTAAGCGG CAGGGTCGGA  
 3571 ACAGGAGAGC GCACGAGGGA GCTTCCAGGG GGAACGCTT GGTATCTTTA TAGTCTGTC GGGTTTCGCC  
 3641 ACCTCTGACT TGAGCGTCTGA TTTTGTGTAT GCTCTGTACG GGGCGGGAGC CTATGGAAAA ACGCCAGCAA  
 3711 CGCGGCTTT TTACGGTTCC TGGCCTTTTG CTGGCCTTTT GCTCACATGT TCTTCTCTGC GTTATCCCCT  
 3781 GATTCTGTGG ATAACCGTAT TACCGCCTTT GAGTGAAGCTG ATACCGCTCG CCGCAGCCGA ACGACCCGAGC  
 3851 GCAGCGAGTC AGTGAAGCGAG GAAGCGGAAG AGCGCCCAAT ACGCAAAACCG CCTCTCCCG CGCGTTGGCC  
 3921 GATTCATTAA TGCAGCTGGC ACGACAGGTT TCCCGAGTGG AAAGCGGCA GTGAGCGCAA CGCAATTAAT  
 3991 GTGAGTTAGC TCACTCATTA GGCACCCGAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA  
 4061 TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCGCG CAATTAACCC  
 4131 TCACTAAAGG GAACAAAAGC TGGGTACCGG GCGCCCTCTC GAGGTCAATC ATATGCTTGA GAAGAGAGTC  
 4201 GGGATAGTCC AAAATAAAC AAAGGTAAAG TTACCTGGTC AAAAGTGAAA ACATCAGTTA AAAGTGGTA  
 4271 TAAGTAAAAAT ATCGGTAAATA AAAGGTGGCC CAAAGTGAAA TTACTCTTT TCTACTATTA TAAAAATTGA  
 4341 GGAATGTTTTG TCGGTACTTT GATACGTCAT TTTGTATGA ATTGGTTTTT AAGTTTATTC GCGATTTGGA  
 4411 AATGCATATC TGTATTTGAG TCGGTTTTTA AGTTCGTGTC TTTTGTAAAT ACAGAGGGAT TTGTATAAGA  
 4481 AATATCTTTA AAAAACCCAT ATGCTAATTT GACATAATTT TTGAGAAAAA TATATATTCA GGCGAATTC  
 4551 ACAATGAACA ATAAATAAGT TAAATAGCT TGCCCGGTT GCAGCGATGG GTATTTTTC TAGTAAATA  
 4621 AAAGATAAAC TTAGACTCAA AACATTTTACA AAAACAACCC CTAAAGTCTT AAAGCCCAA GTGCTATGCA  
 4691 CGATCCATAG CAAGCCGAGC CCAACCCAAC CCACCCCACT GCAGCCCACT GGCAAAATAGT  
 4761 CTCACCCCCC GGCACATATCA CCGTGAAGTG TCCGCAACAC CGCAGCTCTC GCAGCCAAA AAAAAAAG  
 4831 AAAGAAAAAA AAGAAAAAGA AAAACAGCAG GTGGTCCGG GTCTGTGGGG CCGGAAAAAGC GAGGAGGATC

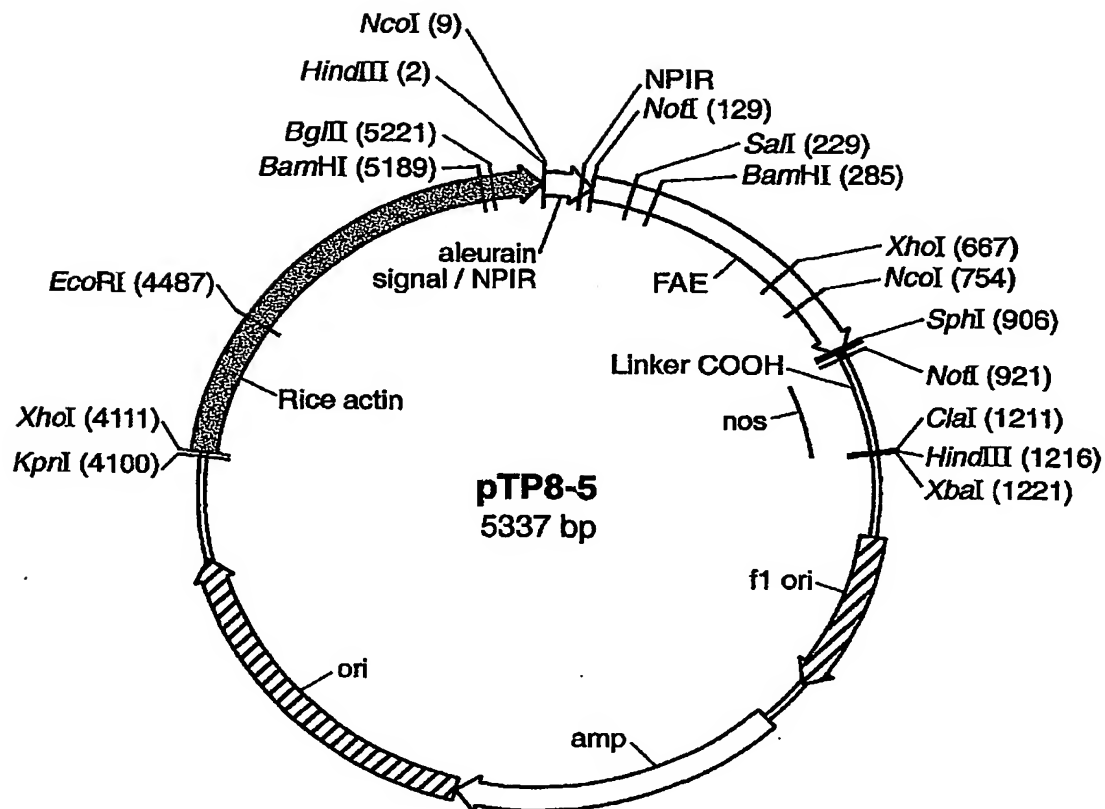
FIG.-35D

```

4901 GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC CGCTTCCAA GAAACGCCCC CCATCGCCAC TATATACATA
4971 CCCCCCCTC TCCTCCCATC CCCCACACCC TACCACCAAC ACCACCAACA CCTCCTCCCC CCTCGCTGCC
5041 GACGACGAG CTCCTCCCC CTCCTCCCTCC GCGCGCGCCG GTAACCAACC GCGCCCTCTC CTCTTCTTTT
5111 CTCCGTTTTT TTTTTCGCT CCGTCTCGAT CTTTGGCCTT GGTAGTTGG GTGGGCGAGA GCGGCTTCGT
                                     BamHI
5181 CGCCACAGATC GGTGCGCGG AGGGCGGGA TCTCGCGGCT GCGTCTCCG GCGTGAATC GGCCCGGATC
                                     BglII
                                     ~~~~~
5251 CTCGCGGGGA ATGGGCTCT CGGATGTAGA TCTTCTTCT TTCTTCTTTT TGTGGTAGAA TTGAATCCC
5321 TCAGCATTTT TCATCGGTAG TTTTCTTTT CATGATTTGT GACAAATGCA GCGTCTGTCG GAGCTTTTTT
5391 GTAGC

```

**FIG. 35E**

**FIG.\_36A**

```

NcoI

HindIII

 M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCCTGGCG TCGCCGTGCT GGCACGGCC GCGTCCGCC
 NotI

 . A S S S F A D S N P I R P V T D R A A A S T .
71 TCGCTCCTC CTCCTCCTC GCGACTCCA ACCGATCCG GCGGTACCC GACCGCGCG CCGCTCCAC
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAGACC TCTACAGCCG TTTAGTCGAA ATGCCACTA TCTCCCAAGC TGCCTACGCC
 Sali

 D L C N I P S F I I K G E K I Y N S Q T D I N G
211 GACCTGTGA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAAACT GACATTAACG
 BamHI

 . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGATCCT CCGGACGAC AGCAGCAAAG AAATAATCAC CGTCTCCGT GGCACGTGTA GTGATACGAA
 . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACACTC GATACTAAT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAAGTA
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGCTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGATC GCTTGTCAA CAGCAGGTTA
 . Q Y P D Y A L T V T G H X L G A S L A L T A .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACKC CCGCGGCC TCCCTGGCGG CACTCAGTGC
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCGAGCTG TCTGGACAT ACGACAACAT CGGCTGTAT ACCTTCGGCG AACCGCGCAG CGGCAATCAG
 XhoI

 A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCGTTCGCGT CGTACATGAA CGATGCCCTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA
 NcoI

 . H A N D G I P N L P P V E Q G Y A H G G V E Y .

```

FIG.\_36B

```

701 CTGATGCCAA CGACGGCATC CCAAACCTGC CCCCCTGGA GCAGGGGTAC GCCCATGGCG GTGTAGATA
 . N B V D P Y S A Q N T F V C T G D E V Q C C E
771 CTGGACGCTT GATCCTTACA GCGCCAGAA CACATTTGC TGCACCTGGG ATGANGTCA GTGCTGTGAG
 sphI
    ~~~~~
    A Q G G Q G V N N A H T T Y F G M T S G A C T W
841 GCCCAGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT
    NotI
    ~~~~~
 . P V A A A *
911 GGCCGGTCGC GGCCGGGTAA CCACTGAAGG ATGAGCTGA AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT
981 AAAGTTTCTT AAGATTGAAT CCTGTGCGG GTCTTGCGAT GATTATCATA TAAATTTCTGT TGAATTTACGT
1051 TAAGCATGTA ATAATTAACTA TGTAAATGCA TACGTTATTT ATGAGATGGG TTTTATATGAT TAGAGTCCCG
1121 CAATTATACA TTTAATACGC GATAGAAAAC AAAATATAGC GCGCAAACTA GGATAAATTA TCGCGCGCGG
 HindIII
    ~~~~~
    ClaI XbaI
    ~~~~~
1191 TGTCTCTCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGCG CGGTGGAGCT CCAATTCGCC CTATAGTGAG
1261 TCGTATTACG CGCGTCACT GGCGTTCGTT TTACACGCTC GTGACTGGGA AAACCTTGGC GTTACCCCAAC
1331 TTAATCGCCT TGCAGCACAT CCCCCTTTTC CCAGCTGCGG TAATAGCGAA GAGGCCCGCA CCGATCGCCC
1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCTGTAGCG GCGCATTAAG CGCGCGCGGT
1471 GTGGTGGTTA CGCGCAGCGT GACCGCTACA CTGCGCAGCG CCTAGCGCC CGCTCCTTTC GCTTTCCTTC
1541 CTTCCTTTCT CGCCACGTTT CCGGCGCTTC CCGGCTAAGC TCTAAATCGG GGGCTCCCTT TAGGGTTCCG
1611 ATTAGTGCT TTACGGCACC TCGACCCCAA AAAACTTGTAT TAGGTGATG GTTCACGTAG TGGGCCCATCG
1681 CCTGATAGA CGGTTTTTCG CCTTTTGACG TTGGAGTCCA CGTCTTTAA TAGTGGACTC TTGTTCCAAA
1751 CTGGAACAAC ACTCAACCTT ATCTCGGTCT ATTTAAGCGG TTTATAAGGG ATTTTCCGA TTTCCGCCCTA
1821 TGGTTAATAA AATGAGCTGA TTTAACAATA AATTTAAGG AATATTAAAC AAATATTAC GCTTACAATT
1891 TAGTGGCAC TTTTCGGGGA AATGTGCGGA AACCCCTAT TTTTATTAT TTTTATTAT TTTCAAAATAT
1961 GTATCCGCTC ATGAGACAAT AACCTGATA AATGCTTCAA TAAATATTGA TTTGCTTCC CACCCAGAAA
2031 CAACATTTCC GTGTGCGCCT TATTTCCCTT TTTGCGGCTT TTTGCTTCC TGTTTTGTCT CACCCAGAAA
2101 CGCTGTGAA AGTAAAGAT GCTGAAGATC AGTTGGGTGC ACAGTGGGT TACATCGAAC TGGATCTCAA
2171 CAGCGTAAG ATCCTTGAGA GTTTTCGCC CGAAGAACGT TTTCCAATGA TGAGCACTTT TAAAGTTCTG
2241 CTATGTGGCG CGGTATTATC CCGTATTGAC GCGGGGCAAG AGCAACTCGG TCGCCGCATA CACTATTCTC
2311 AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAGCA TCTTACGGAT GGCATGACAG TAAGGAATT

```

FIG. 36C



**FIG. 36D**

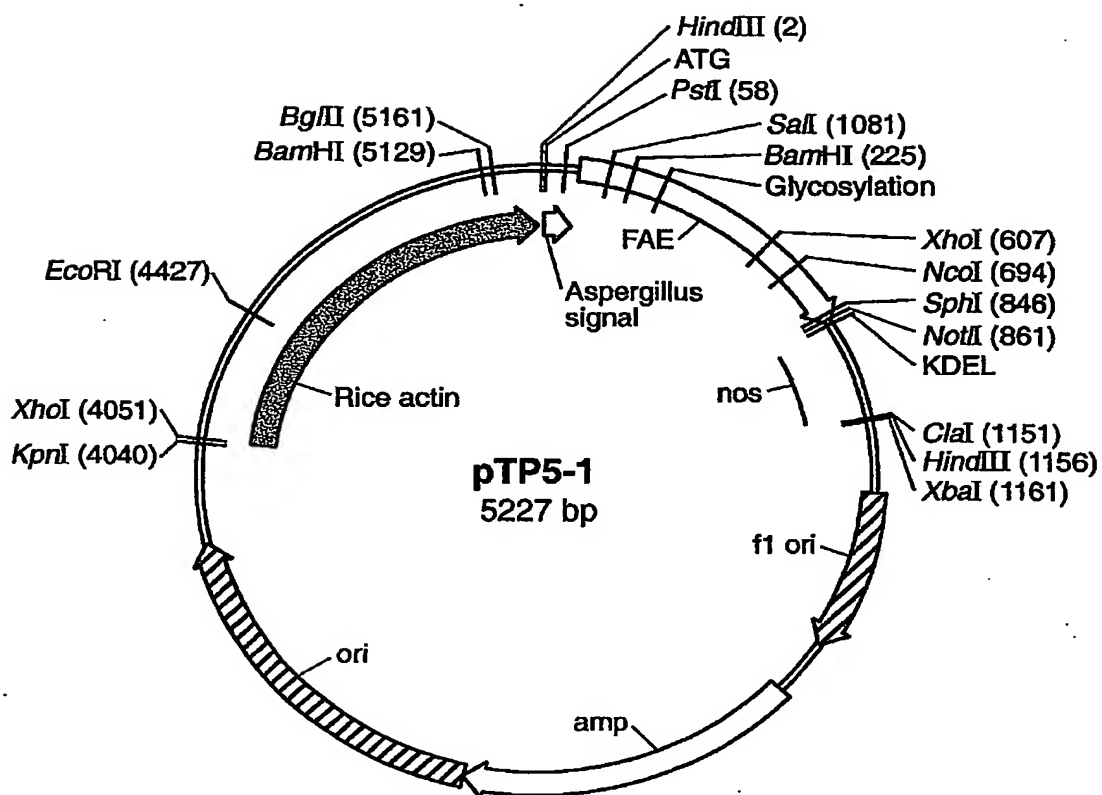
```

4551 TCTAGTAAAA TAAAGATAA ACTTAGACTC AAACATTTA CAAAACAAC CCTAAAGTC CTAAGGCCA
4621 AAGTGCTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA ACCCAACCA ACCCACCCA GTGCAGCCAA
4691 CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTGAGT TGTCCGAGC ACCGACGTC TCGCAGCCAA
4761 AAAAAAAA AGAAGAAAA TCGGAGCAG CGACGAGCC CGGCCCTCCC TCCGTTCCA AAGAAACGCC CCCCATCGCC
4831 GCGAGGAGGA TCGGAGCAG TCTCTCTCCA TCTCTCTCCA CCTACCACTA CACCAACGCC CACCTCCTCC
4901 ACTATATACA TACCCCTCCC AGCTCCTCCC CCTTCCCTCC CCGCGCGCG CCGTAACCCAC CCGGCCCTC
4971 CCCCTCGCTG CCGGACGACG TTTCTCCGTTT TTTCTCTCGT CTGCGTCTCG ATCTTTGGCC TTGGTAGTTT GGGTGGGCGA
5041 TCCCTCTTCT TTTCTCCGTTT TTTCTCTCGT TTTCTCTCGT GATCTCGCG GATCTCGCG CTGGCGTCTC CGGGCGTGA
5111 GAGCGGCTTC GTCGCCCCAG TCGGTGCGCG GGAGGGGCGG GATCTCGCG GATCTCGCG CTGGCGTCTC CGGGCGTGA

 BamHI
                                     ~~~~~
5181 TCGGCCCGGA TCCTCGCGGG GAATGGGGCT CTCGGATGA GATCTCTTT CTTTCTTCTT TTTGTGGTAG
5251 AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTCTTT TTCAATGATT GTGACAAATG CAGCCTCGTG
5321 CCGAGCTTTT TTGTAGC

```

**FIG. 36E**

**FIG.\_37A**

```

HindIII
~~~~~
M K Q F S A K H V L A V V V T A G H A L A
1 AAGCTTAACA TGAAGCAGTT CTCCGCCAA CAGTCTCTCG CAGTTGTGCT GACTGCAGGG CAGCCCTTAG
. A S T Q G I S E D L Y S R L V E M A T I S Q A .
71 CAGCCTCTAC GCAAGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC
SalI
~~~~~
. A Y A D L C N I P S T I I K G E K I Y N S Q T
141 TGCCTACGCC GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAAACT
BamHI
~~~~~
D I N G W I L R D D S S K E I I T V F R G T G S
211 GACATTAACG GATGGATCCT CCGCGACGAC AGCAGCAAAG AATAATCAC CGTCTTCCGT GGCACCTGGTA
. D T N L Q L D T N Y T L T P F D T L P Q C N G .
281 GTGATACGAA TCTACAACATC GATACTAACT ACACCTCTAC GCCTTTGAC ACCCTACCAC AATGCAACGG
. C E V H G G Y Y I G W V S V Q D Q V E S L V K
351 TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAAA
Q Q V S Q Y P D Y A L T V T G H X L G A S L A A
421 CAGCAGTTA GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACAC CCTCGGCGCC TCCCTGGCGG
. L T A A Q L S A T Y D N I R L Y T F G E P R S .
491 CACTCACTGC CGCCAGCTG TCTGCGACAT ACGACAACAT CCGCTGTATC ACCTTCGGCG AACCGCGCAG
XhoI
~~~~~
. G N Q A F A S Y M N D A F Q A S S P D T T Q Y
561 CGGCAATCAG GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCTCGA GCCCAGATAC GACGCAGTAT
NcoI
~~~~~
F R V T H A N D G I P N L P P V E Q G Y A H G G
631 TTCCGGGTCA CTCATGCCAA CGACGGCATC CCAACCTGC CCCGGGTGA GCAGGGGTAC GCCCATGGCG
. V E Y W S V D P Y S A Q N T F V C T G D E V Q .
701 GTGTAGATA CTGGAGCGTT GATCCTTACA GCGCCCAAGAA CACATTGTC TGCACCTGGG ATGAAGTCCA
. C C E A Q G G Q G V N N A H T T Y F G M T S G
771 GTGCTGTGAG GCCCAGGCG GACAGGGTGT GAATAATGCG CACAGCACTT ATTTTGGGAT GACGAGCGGC

```

FIG.-37B

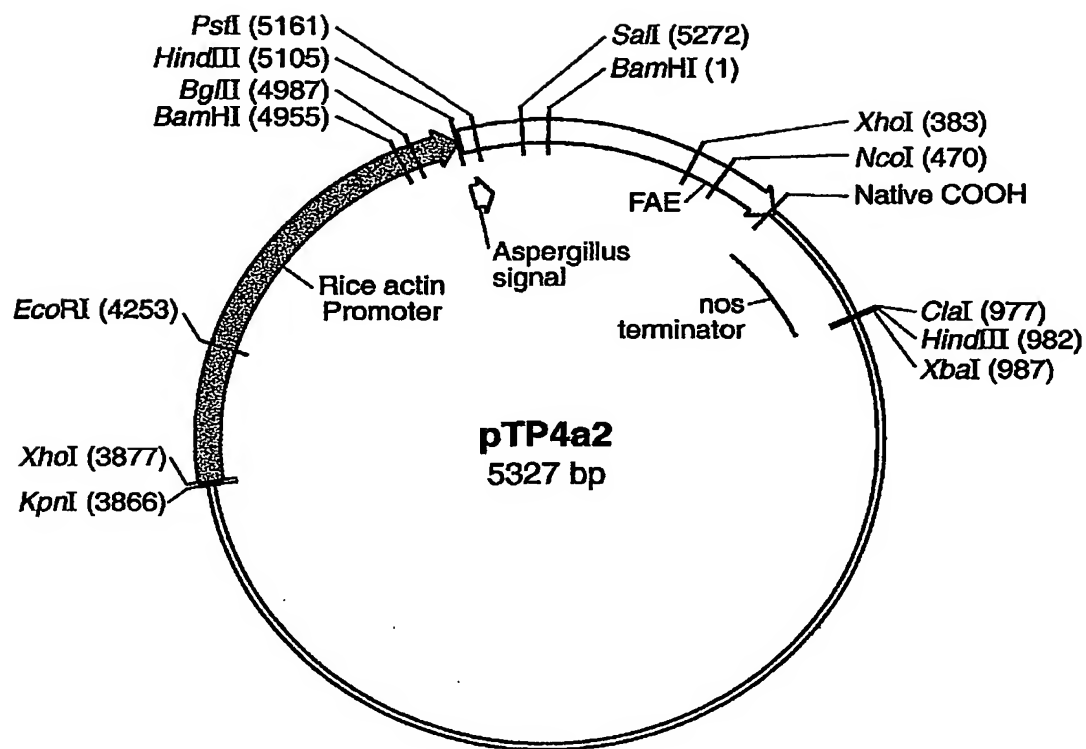
| SphI                              |             |            |            |             |            |            |             |  |  |  |  |  |  |  |  |
|-----------------------------------|-------------|------------|------------|-------------|------------|------------|-------------|--|--|--|--|--|--|--|--|
| NotI                              |             |            |            |             |            |            |             |  |  |  |  |  |  |  |  |
| A C T W P V A A A E P L K D E L * |             |            |            |             |            |            |             |  |  |  |  |  |  |  |  |
| 841                               | GCATGCACCT  | GGCCGGTCCG | GGCCGGGCGA | CCACTGAAG   | ATGAGCTGTA | AAGAGCAGA  | TCGTTCAAAC  |  |  |  |  |  |  |  |  |
| 911                               | ATTGGCAAT   | AAAGTTCTT  | AGATTGAAT  | CCTGTTGCCG  | GTCTTGCGAT | GATTATCAT  | TAATTTCTGT  |  |  |  |  |  |  |  |  |
| 981                               | TGAATTACGT  | TAAGCATGTA | ATAATTACA  | TGTAATGCA   | GACGTTATTT | ATGAGATGGG | TTTTTATGAT  |  |  |  |  |  |  |  |  |
| 1051                              | TAGAGTCCCG  | CAATTATACA | TTTAATACGC | GATAGAAAAC  | AAAATATAGC | GCACAACATA | GGATAAATTA  |  |  |  |  |  |  |  |  |
| HindIII                           |             |            |            |             |            |            |             |  |  |  |  |  |  |  |  |
| ClaI                              |             |            |            |             |            |            |             |  |  |  |  |  |  |  |  |
| XbaI                              |             |            |            |             |            |            |             |  |  |  |  |  |  |  |  |
| 1121                              | TCGCGCGCGG  | TGTCATCTAT | GTTACTAGAT | CGATAAGCTT  | CTAGAGCGGC | CGGTGGAGCT | CCAAATTCGCC |  |  |  |  |  |  |  |  |
| 1191                              | CTATAGTGAG  | TCGTATFACG | CGCGTCACT  | GGCCGTCTGT  | TTACAACGTC | GTGACTGGGA | AAACCCCTGGC |  |  |  |  |  |  |  |  |
| 1261                              | GTTACCCCAAC | TTAATCGCCT | TGCAGCACAT | CCCCCTTTCG  | CCAGCTGGCG | TAATAGCGAA | GAGGCCCGCA  |  |  |  |  |  |  |  |  |
| 1331                              | CCGATCGCCC  | TTCCCAACAG | TTGCGCAGCC | TGAATGGCGA  | ATGGGACGCG | CCCTGTAGCG | GCGCATTAAG  |  |  |  |  |  |  |  |  |
| 1401                              | CGCGGCGGGT  | GTGGTGGTTA | CGCGCAGCGT | GACCGCTACA  | CTTGCCAGCG | CCCTAGCGCC | CGCTCCTTTC  |  |  |  |  |  |  |  |  |
| 1471                              | GCTTCTTCC   | CTTCCCTTCT | CGCCACGTTT | GCCGGCTTTC  | CCCGTCAAGC | TCTAAATCGG | GCGCTCCCTT  |  |  |  |  |  |  |  |  |
| 1541                              | TAGGTTCCG   | ATTAGTGCT  | TTACGGCACC | TCGACCCCAA  | AAAACTTGAT | TAGGGTGATG | GTTTACGTTAG |  |  |  |  |  |  |  |  |
| 1611                              | TGGGCCATCG  | CCCTGATAGA | CGGTTTTCG  | CCCTTTGACG  | TTGGAGTCCA | CGTTCTTTAA | TAGTGGACTC  |  |  |  |  |  |  |  |  |
| 1681                              | TTTGTCCAAA  | CTGGAAACA  | ACTCAACCTT | ATCTCGGTCT  | ATTCCTTTGA | TTTATAAGGG | ATTTTGCCGA  |  |  |  |  |  |  |  |  |
| 1751                              | TTTCGGCCFA  | TTGGTTAAAA | AATGAGCTGA | TTTAAACAAA  | ATTTAACCGG | AAATTTAACA | AAATATTAAAC |  |  |  |  |  |  |  |  |
| 1821                              | GCTTACAAAT  | TAGGTGGCAC | TTTTCGGGGA | AATGTGCGCG  | GAACCCCTAT | TTGTTTATTT | TTCTAAATAC  |  |  |  |  |  |  |  |  |
| 1891                              | ATTCAAAAT   | GTATCCGCTC | ATGAGACAAT | AACCCCTGATA | AATGCTTCAA | TAATATTGAA | AAAGGAAGAG  |  |  |  |  |  |  |  |  |
| 1961                              | TATGAGTATT  | CAACATTTCC | GTGTCGCCCT | TATTCCTTTT  | TTTGGGCGAT | TTTGCCTTCC | TGTTTTTGCT  |  |  |  |  |  |  |  |  |
| 2031                              | CACCCAGAAA  | CGCTGGTGAA | AGTAAAGAT  | GCTGAAGATC  | AGTTGGGTGC | ACGAGTGGGT | TACATCGAAC  |  |  |  |  |  |  |  |  |
| 2101                              | TGGATCTCAA  | CAGCGGTAAG | ATCCTTGAGA | GTTTTTCGCC  | CGAAGAACGT | TTTCCAATGA | TGAGCACTTT  |  |  |  |  |  |  |  |  |
| 2171                              | TAAAGTCTG   | CTATGTGGCG | CGGTATTATC | CCGTATTGAC  | GCCGGSCAAG | AGCAACTCGG | TCGCCCCATA  |  |  |  |  |  |  |  |  |
| 2241                              | CACATATTCY  | AGATGACTT  | GGTTGAGTAC | TCACCAGTCA  | CAGAAAAGCA | TCTTACGGAT | GGCATGACAG  |  |  |  |  |  |  |  |  |
| 2311                              | TAAGAGATT   | ATGCAGTGT  | GCCATAACCA | TGAGTGATAA  | CACGTGGGCC | AACTTACTTC | TGACAAACGAT |  |  |  |  |  |  |  |  |
| 2381                              | CGAGGAGCCG  | AAGGAGCTAA | CCGCTTTTIT | GCACAACATG  | GGGGATCATG | TAACTCGCCT | TGATCGTTGG  |  |  |  |  |  |  |  |  |
| 2451                              | GAACCGGAGC  | TGAATGAAGC | CATACCACAA | GACGAGCGTG  | ACACCACGAT | GCCTGTAGCA | ATGGCAACAA  |  |  |  |  |  |  |  |  |
| 2521                              | CGTTGGCGAA  | ACTATTAACT | GGCGAACTAC | TTACTCTAGC  | TTCCGGGCAA | CAATTAATAG | ACTGGATGGA  |  |  |  |  |  |  |  |  |
| 2591                              | GGCGGATAAA  | GTTCAGGAC  | CACCTCTGCG | CTCGGCCCTT  | CGGCTGGCT  | GGTTTATTGC | TGATAAATCT  |  |  |  |  |  |  |  |  |
| 2661                              | GGAGCCCGTG  | AGCGTGGGTC | TCGCGGTATC | ATTGCAGCAC  | TGGGGCCAGA | TGGTTRAGCC | TCCCCTATCG  |  |  |  |  |  |  |  |  |

FIG.-37C

|      |             |            |            |             |             |             |             |
|------|-------------|------------|------------|-------------|-------------|-------------|-------------|
| 2731 | TAGTTATCTA  | CACGACGGGG | AGTCAGGCAA | CTATGGATGA  | ACGAATAGA   | CAGATCGCTG  | AGATAGGTGC  |
| 2801 | CTCAGTGATT  | AAGCATGGT  | AACTGTGAGA | CCAAAGTTTAC | TCATATATAC  | TTTAGATTGA  | TTTAAACCTT  |
| 2871 | CATTTTAAAT  | TTAAAGGAT  | CTAGGTGAAG | ATCCTTTTGG  | ATAATCTCAT  | GACCAAAATC  | CCTTAAACGTG |
| 2941 | AGTTTTCGTT  | CCACTGAGCG | TCAGACCCCG | TAGAAAAGAT  | CAAGGATCT   | TCTTGAGATC  | CTTTTTCCT   |
| 3011 | GCCTGTAATC  | TGCTGCTTGC | AAACAATAAA | ACCACCGCTA  | CCAGCGGTGG  | TTTGTGTTGC  | GGATCAAGAG  |
| 3081 | CTACCAACTC  | TTTTTCCGAA | GGTAAGTGG  | TTTACACTGC  | CAGAGATACC  | AAATACTGTC  | CTTCTAGTGT  |
| 3151 | AGCGTAGTGT  | AGGCCACAC  | TTCAAGAACT | CTGTAGCACC  | GCCTACATAC  | CTCGCTCTGC  | TAATCCTGTT  |
| 3221 | ACCAGTGGCT  | GCTGCCAGTG | GCATATAGTC | TGTCTTACC   | GGTGTGGACT  | CAAGACGATA  | GTTACCCGAT  |
| 3291 | AAGGCGCAGC  | GGTCGGGCTG | AACGGGGGGT | TGCTGCACAC  | AGCCAGCTT   | GGAGCGAAGG  | ACCTACACCG  |
| 3361 | AACTGAGATA  | CCTACAGCGT | GAGCTATGAG | AAAGCGCCAC  | GCTTCCCGAA  | GGAGAAAAGG  | CGGACAGGTA  |
| 3431 | TCCGTAAGC   | GGCAGGTCG  | GAACAGGAGA | CGGCACGAGG  | GAGCTTCCAG  | GGGAAAACGC  | CTGGTATCTT  |
| 3501 | TATAGTCCCTG | TCGGGTTCG  | CCACCTCTGA | CTTGAGCGTC  | GATTTTGTG   | ATGCTCGTCA  | GGGGGGCGGA  |
| 3571 | GCCTATGGAA  | AAACGCCAGC | AACGCGGCTT | TTTTACGGTT  | CCTGGCCTTT  | TGCTGGCCTT  | TTGCTCACAT  |
| 3641 | GTTCTTTCCT  | GCCTTATCCC | CTGATTCCTG | GGATAACCGT  | ATTACCGCCT  | TTGAGTGAGC  | TGATACCGCT  |
| 3711 | CGCCGCAGCC  | GAACGACCGA | CGCAGCGAG  | TCAGTGAGCG  | AGGAAGCGGA  | AGAGCGCCCA  | ATACGCAAAC  |
| 3781 | CGCTCTCCC   | CGCGGCTTGG | CCGATTCAAT | AATGCAAGTC  | GCACGACAGG  | TTTCCCGACT  | GGAAAGCGGG  |
| 3851 | CAGTGAGCGC  | AACGCAATTA | ATGTGAGTTA | GCTCACTCAT  | TAGGCACCCC  | AGGCTTTACA  | CTTTATGCTT  |
| 3921 | CCGGCTCGTA  | TGTTGTGTGG | AATTGTGAGC | GGATAACAAT  | TTACACACAGG | AAACAGCTAT  | GACCATGATT  |
|      |             |            |            |             | KpnI        |             | XhoI        |
| 3991 | ACGCCAAGCG  | CGCAATTAAC | CCTCACTAAA | GGGAACAATA  | GCTGGGTACC  | GGGCCCCCCC  | TCGAGGTCTAT |
| 4061 | TCATATGCTT  | GAGAAGAGAG | TCGGGATAGT | CCAAATATAA  | ACAAAGGTAA  | GATTACCTGG  | TCAAAAGTGA  |
| 4131 | AAACATCAGT  | TAAAGGTGG  | TATAAGTAAA | ATATCGGTAA  | TAAAAGGTGG  | CCCAAGTGA   | AATTTACTCT  |
| 4201 | TTTCTACTAT  | TATAAAAAT  | GAGGATGTTT | TGTCGGTACT  | TTGATACGTC  | ATTTTGTAT   | GAAATTGGTTT |
| 4271 | TTAAGTTTAT  | TCGCGATTGG | GAAATGCATA | TCTGTATTTG  | AGTCGGTTT   | TAAATTCTGT  | GCTTTTGTAA  |
| 4341 | ATACAGAGGG  | ATTGTATATA | GAAATATCTT | TAAAAAACCC  | ATATGCTAAT  | TTGACATAAT  | TTTTTGAGAA  |
|      |             |            |            |             | EcoRI       |             |             |
| 4411 | AATATATATT  | CAGGCGAATT | CCACAATGAA | CAATAATAAG  | ATTAATAAG   | CTTGCCCCCG  | TTTGCAGCGAT |
| 4481 | GGGTATTTT   | TCTAGTAAAA | TAAAGATATA | ACTTAGACTC  | AAAAATTTTA  | CAAAAACAAAC | CCCTAAAGTC  |
| 4551 | CTAAAGCCCA  | AAGTGCTATG | CACGATCCAT | AGCAAGCCCA  | GCCCAACCCA  | ACCAACCCA   | ACCCACCCCA  |
| 4621 | GTGCAGCCAA  | CTGGCAATA  | GTCTCCACCC | CCGGCACTAT  | CACCGTGAAT  | TGTCCGCACC  | ACCGCACGTC  |
| 4691 | TCGCAGCCAA  | AAAAAATAA  | AGAAAGATAA | AAAAAATAA   | GAAAAACAGC  | AGGTGGGTCC  | GGGTCTGTGG  |

FIG. 37D

**FIG. 37E**

**FIG.\_38A**



BamHI  
 .....  
 . I L R D D S S K E I I T V F R G T G S D T N L  
 1 GATCCTCCGC GACGACAGCA GCAAAGAAT AATCACCGTC TTCGTTGGCA CTGGTAGTGA TACGAATCTA  
 Q L D T N Y T L T P F D T L P Q C N G C E V H G  
 71 CAACTCGATA CTAACATACAC CCTCAGCCTT TCGACACCC TACACAATG CAACGGTTGT GAAGTACACG  
 . G Y Y I G W V S V Q D Q V E S L V K Q Q V S Q .  
 141 GTGGATATTA TATTGGATGG GTCTCCGTCC AGGACCAAGT CGAGTCGCTT GTCAACACAGC AGGTTAGCCA  
 . Y P D Y A L T V T G H X L G A S L A A L T A A  
 211 GTATCCGGAC TACGCGCTGA CCGTGACCG CCACKCCCTC GCGCCTCC TGGCGGCACT CACTGCCGCC  
 Q L S A T Y D N I R L Y T F G E P R S G N Q A F  
 281 CAGCTGCTG CGACATACGA CAACATCCGC CTGTACACCT TCGGCGAACC GCGCAGCGGC AATCAGGCGCT  
 XhoI  
 .....  
 . A S Y M N D A F Q A S S P D T T Q Y F R V T H .  
 351 TCGCGTCGTA CATGAACGAT GCCTTCCAAG CCTCGAGCCC AGATACGACG CAGTATTTC CCGTCACTCA  
 NcoI  
 .....  
 . A N D G I P N L P P V E Q G Y A H G G V E Y W  
 421 TGCCAAACGAC GGCATCCCAA ACCTGCCCCC GGTGGAGCAG GGTACGCC C ATGGCGGTGT AGAGTACTGG  
 S V D P Y S A Q N T F V C T G D E V Q C C E A Q  
 491 AGCGTTGATC CTTACAGCGC CCAGAACACA TTGTGCTGCA CTGGGGATGA AGTCAGTGC TGTGAGGCCC  
 . G G Q G V N N A H T T Y F G M T S G A C T W \* .  
 561 AGGCGGACA GGTGTGAAT AATGCGACA CGACTTATT TGGGATGACG AGCGGAGCCT GTACATGGTG  
 . \*  
 631 ATCAGTCATT TCAGCCTCCC CGAGTGATCC AGGAAGATG GATGTCCTGG AGAGGGGGCC GCGTAACCCAC  
 701 TGAAGGATGA GCTGTAAAGA AGCAGATCGT TCAACACATT GSCAATAAAG TTCTCTAAGA TTGAATCCTG  
 771 TTGCCGGTCT TCGATGATAT ATCATATAAT TTCTGTGAA TTACGTTAAG CATGTAATAA TTAACATGTA  
 841 ATGCATGACG TTATTATGA GATGGGTTTT TATGATTAGA GTCCCGCAAT TATACATTTA ATACGCGATA  
 ClaI  
 911 GAAACAAAA TATAGCGCGC AAACCTAGGAT AAATTATCGC GCGCGGTGTC ATCTATGTTA CTAGATCGAT  
 XbaI  
 .....  
 HindIII  
 .....

FIG.-38B

981 AAGCTTCTAG AGCGGCCGGT GGAGCTCCAA TTCGCCCTAT AGTGATCGT ATTAGCGCG CTCACCTGGCC  
 1051 GTCGTTTAC AACGTCGTGA CTGGGAAAC CTTGGCGTTA CCCAACTAA TCGCCTTGCA GCACATCCCC  
 1121 CTTTCGCCAG CTGGCGTAAT AGCGAAGAG CCAGCACCGA TCGCCTTCC CAACAGTTGC GCAGCCTGAA  
 1191 TGGCGAATGG GACGCGCCCT GTAGCGGCG ATTAAGCGCG GTGGGTGTGG CGGGGTGTGG CAGCGTGACC  
 1261 GCTACACTTG CCAGCGCCCT AGCGCCCGCT CTTTCGCTT CTTTCGCTT TCTTCCTTC CAGTTGCGCG  
 1331 GCTTCCCGG TCAAGCTCTA AATCGGGGG TCCCTTTAGG GTTCGGATTT AGTGCTTTAC GGCACCTCGA  
 1401 CCCCAGGAAA CTTGATAGG GTGATGGTTC AGGTAGTGG CCATCGCCCT GATAGACGGT TTTTCGCCCT  
 1471 TGACGCTTGG AGTCCACGTT CTTTAAATAG GCACTCTTGT TCCAAACTG AACCAACTC AACCTATCT  
 1541 CGGCTATTC TTTTGATT TAAGGATTT TCGCGATTG GGCCTATTGG GTTAAAAATG AGCTGATTTA  
 1611 ACAGAAATTT AACGCGAATT TTAACAAAT ATTAACGCTT ACAATTTAGG TGGCACTTTT CGGGGAAATG  
 1681 TGGCGGGAAC CCTATTGT TTATTTTCT AAATACATC AAATATGTAT CCGCTCATGA GACAAATAAC  
 1751 CTGATAATG CTTCAATAAT ATTGAAAAG GAAGATATG AGTATTC AAC ATTTCCGTGT CGCCCTTATT  
 1821 CTTTCTTTG CGGCATTTT CTTCTCTGT TTTGCTCACC CAGAAACGCT GGTGAAAATA AAAGATGCTG  
 1891 AAGATCAGTT GGTGACAGA GTGGGTIACA TCGAACTGGA TCTCAACAGC GGTAAAGATCC TTGAGAGTTT  
 1961 TCGCCCGAA GAACGTTTTC CAATGATGAG CACTTTTAAA GTTCTGTAT GTGGCGCGGT ATTATCCCGT  
 2031 ATTGACGCC GGCAGAGCA ACTCGGTGCG CGCATACAT ATTCTCAGAA AGTCTTGTT GAGTACTCAC  
 2101 CAGTACAGA AAAGCATCTT ACGGATGGCA TGACAGTAA AGAATATATG AGTCTGCCA TAACCATGAG  
 2171 TGATAACACT CGCGCAACT TACTTCTGAC AACGATCGA GACCGAAGG AGCTAACCG TTTTTCGAC  
 2241 AACATGGGG ATCATGTAAC TCGCCTTGAT CGTTGGGAA CGGAGCTGAA TGAAGCCATA CCAACCGACG  
 2311 AGCGTGACAC CACGATGCC CTAGCAATGG CAACAACGTT GCGCAAACTA TTAACGTGGC AACTACTTAC  
 2381 TCTAGCTTCC CGGCAACAAT TAATAGACTG GATGGAGCG GATAAAGTTG CAGGACCACT TCTGCGCTCG  
 2451 GCCCTTCGG CTGGCTGGTT TATTGCTGAT AAATCTGGAG CCGGTGAGCG TGGGTCTCG GGTATCATTG  
 2521 CAGCACTGG GCCAGATGGT AAGCCCTCCC GTATCGTAGT TATCTACAG ACGGGAGTC AGGCAACTAT  
 2591 GGATGAACGA AATAGACAGA TCGCTGAGT AGGTGCCCTA CTGATTAAGC ATTTGTAAC GTGAGATCC  
 2661 GTTACTCAT ATATACTTTA GATTGATT TAACCTCAT AAACCTCAT TTTAATTTAA AAGGATCTAG GTGAAGATCC  
 2731 TTTTGTATAA TCTCATGACC AAAATCCCTT AACGTGAGTT TTCGTTCCAC TGAGCGTCA ACCCGTAGA  
 2801 AAAGATCAA GGATCTTCTT GAGATCCCTT TTTTCTGCGC GATACTGCT GCTTGCAAC AAAAAACCA  
 2871 CCGCTACCAG CGGTGGTTTG TTTGCCGGAT CAAGAGCTAC CAACTCTTTT TCCGAAGGTA ACTGGCTTCA  
 2941 GCAGAGCGCA GATACCAAT ACTGTCTTTC TAGTGTAGCC GTAGTTAGGC CACCACCTCA AGAATCTGT  
 3011 AGCAGCGCT ACATACCTCG CTCTGCTAAT CTTGTACCA GTGGCTGCTG CCAGTGGCGA TAAGTCTGT  
 3081 CTTACCGGGT TGGACTCAAG ACGATAGTTA CCGGATAAG CCAGCGGTC CGAGCTGAAC GGGGTTCGT  
 3151 GCACACAGCC CAGCTTGGAG CGAACGACCT ACACCGAAT GAGATACCTA CAGCGTAGC TATGAGAAAG  
 3221 CGCCACGCTT CCCGAAGGGA GAAAGCGGA CAGGTATCCG GTAAAGCGCA GGTTCGGAAC AGGAGAGCGC  
 3291 ACAGAGGAGC TTCCAGGGGG AAACGCTGG TATCTTATA GTCTGTCTG GTTTCGCCAC CTCTGACTTG  
 3361 AGCGTCAATT TTTGTGATGC TCGTCAGGGG GCGCGAGCCCT ATGAAAAAC GCCAGCAACG CGGCCCTTTT

FIG.-38C

```

3431 ACGGTTCCCTG GCCTTTTGCT GGCCTTTTGC TCACATGTTT TTTCTCTGCGT TATCCCTCTGA TTCCTGTGGAT
3501 AACCGTATTA CCGCCTTTGA GTGAGCTGAT ACCGTCGCGC GCAGCCGGAAC GACCGAGCGC AGCGAGTCAG
3571 TGAGCGAGGA AGCGGAAGAG CGCCCAATAC GCAACCGCC TCTCCCCCGG CGTTGGCCGA TTCATTATATG
3641 CAGCTGGCAC GACAGGTTTC CCGACTGGAA AGCGGGCAGT GAGCGCAACG CAATTATATG GAGTTAGCTC
3711 ACTCATTAGG CACCCAGGC TTACACATTT ATGCTTCGG CTGCTATGTT GTGTGGAATT GTGAGCGGAT
3781 AACAAATTCA CACAGGAAC AGCTATGACC ATGATATACG CARGCGCGCA ATTAACCCCTC ACTAAAGGGA

 KpnI XhoI
          ~~~~~
3851  AAAAAAGCTG GGTACCGGGC CCCCCCTCGA GGTCAATTCAT ATGCTTGAGA AGAGAGTCGG GATAGTCCAA
3921  AATAAAACAA AGGTAAGATT ACCTGGTCAA AAGTGAAAC ATCAGTTAAA AGGTGGTATA AGTAAATAT
3991  CGGTAATAAA AGGTGGCCCA AAGTGAATTT TACTCTTTTC TACTATTATA AAAATTGAGG ATGTTTGTTC
4061  GGTACTTTGA TACGTCAATTT TTGTATGAAT TGGTTTTTAA GTTTATTCGC GATTTGGAAT TGCATATCTG
4131  TATTGAGTC GGTTTTAAAG TTCGTTTGCTT TTGTAATATC AGAGGGATTT GTATAAGAAA TATCTTTTAA

          EcoRI
          ~~~~~
4201 AAACCCATAT GCTAATTTGA CATAATTTT GAGAAATAA TATATTCAGG CGAATTCAC AATGAACAT
4271 AATAAGATTA AATAGCTTG CCCCCTGTTG AGCGATGGGT ATTTTTCCTA GTAAATATAA AGATAAACAT
4341 AGACTCAAAA CATTTACAAA AACCAACCC AACCAACCC ACCCACTGC AGCCAAAGT GCTATGCACG ATCCCTAGCA
4411 AGCCAGCCC AACCACACC AACCACACC AACCACACC AACCACACC AACCACACC AACCACACC AACCACACC
4481 CACTATCACG GTGAGTTGTC CGCACCACCG CACGTCGCG CAGTCTGCG AGCCAAATAA AGAATAAATA
4551 GAAAAAGAAA AACAGCAGG GGTCCCGGGT CGTGGGGGCC GAAAAAGCGA GGAGGATCGC GAGCAGCGAC
4621 GAGGCCCGGC CCTCCCTCCG CTTCACAAGA AACGCCCCC ATCGCCACTA TATACATACC CCCCCCTCTC
4691 CTCCCATCCC CCAACCCCTA CCACCAACC ACACCAACC CACTCCCTCC TCGCTGCCCG ACGACGAGCT
4761 CCTCCCCCTT CCCCCCTCCG CGCCGCGCGT AACCAACCG CCCCCTCTCT CTTCTCTTCT CCGTTTTTFT
4831 TTTCTGTCCTG GTCTCGATCT TTGGCCCTTG TTGTTTGGGT GGGCGAGAGC GGCTTCGTCT CCCAGATCGG

 BamHI
          ~~~~~
4901  TGCGCGGGAG GGGCGGGATC TCGCGGCTGG CGTCTCCGG CGTGAGTCGG CCCGGATCCT CGCGGGGAAT

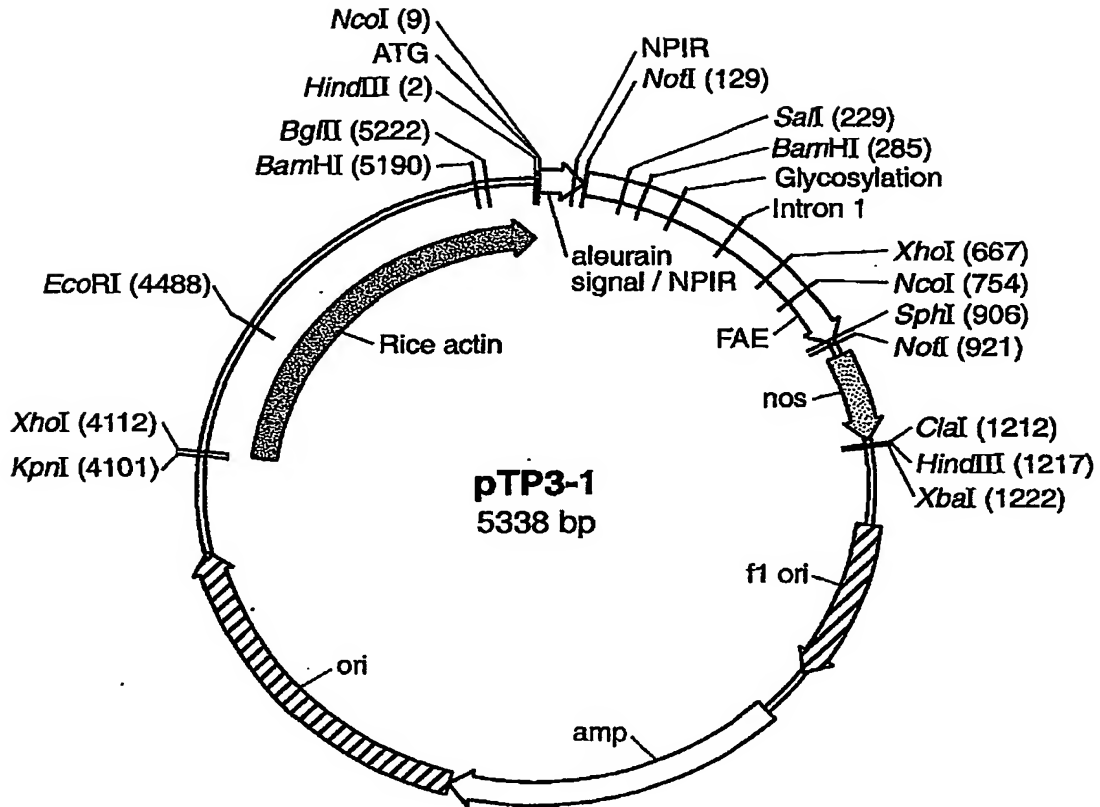
          BglII
          ~~~~~
4971 GGGGCTCTCG GATGTAGATC TTCTTTCTTT TTCTTTCTTT TGTGTAGAAT TGAATCCCTC AGCATTTGTT
 HindIII
          ~~~~~
5041  ATCGGTAGTT TTTCTTTTCA TGATTGTGA CAAATGCAG CTCGTGCGGA GCTTTTTTGT AGCAAGCTTA

```

FIG. 38D

PstI  
 M K Q F S A K H V L A V V V T A G H A L A A S .  
 5111 ACATGAAGCA GTTCTCCGCC AACACGTC TCCAGTTGT GGTGACTGCA GGGCAGCCT TAGCAGCCTC  
 . T Q G I S E D L Y S R L V E M A T I S Q A A Y  
 5181 TAGCAAGGC ATCTCCGAG ACCTCTACAG CCGTTAGTC GAAATGGCCA CTATCTCCA AGCTGCTAC  
 Sali  
 A D L C N I P S T I I K G E K I Y N S Q T D I N  
 5251 GCCGACCTGT GCAACATTCC GTCGACTATT ATCAAGGAG AGAAATTTA CAATTCTCAA ACTGACATTA  
 B  
 . G W  
 5321 ACGGATG

**FIG.-38E**

**FIG. 39A**

NcoI  
 ~~~~~  
 HindIII
 ~~~~~  
 1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGGC TGGCCGTGCT GGCACGGCC GCGTGGCG  
 W A H A R V L L L A L A V L A T A A V A V  
 NotI  
 ~~~~~  
 71 TCGCTCCTC CTCTCCTC GCGACTCA ACCGATCC GCGGTACC GACGGCGG CCGCTCCAC
 . A S S S F A D S N P I R P V T D R A A A S T
 . Q G I S E D L Y S R L V E M A T I S Q A A Y A
 141 GCAGGGCATC TCGAAGACC TCTACAGCG TTAGTCGAA ATGCCACTA TCTCCCAAGC TGCCTACGCC
 Sali
 ~~~~~  
 211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTACG  
 D L C N I P S T I I K G E K I Y N S Q T D I N G  
 BamHI  
 ~~~~~  
 281 GATGGATCCT CCGGACGAC AGCAGCAAG AATAATCAC CGTCTCCGT GGCACCTGGTA GTGATACGAA
 . W I L R D D S S K E I I T V F R G T G S D T N
 Glycosylation
 ~~~~~  
 351 TCTACAACTC GATACTAAT ACACCTCAC GCCTTCGAC ACCCTACCAC AATGCAACG TTGTGAAGTA  
 L Q L D T N Y T L T P F D T L P Q C N G C E V  
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S  
 421 CACGGTGGAT ATTATATGG ATGGTCTCC GTCCAGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
 . Q Y P D Y A L T V T G H X L G A S L A A L T A  
 491 GCCAGTATCC GGAATACGG CTGACCGTGA CCGGCCACKC CCTCGGGGCC TCCCTGGCGG CACTCACTGC  
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q  
 561 CGCCAGCTG TCTGGACAT ACGACAACAT CCGCTGTAC ACCTCGGCG AACCGCGCAG CGGCAATCAG

FIG. 39B

XhoI  
 ~~~~~  
 A F A S Y M N D A F Q A S S P D T T Q Y F R V T
 631 GCCTTCGCGT CGTACATGAA CGATGCCCTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA

 NcoI
 ~~~~~  
 . H A N D G I P N L P P V E Q G Y A H G G V E Y .  
 701 CTCATGCCAA CGACGGCATC CCAACCTGC CCCCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
 . W S V D P Y S A Q N T F V C T G D E V Q C C E  
 771 CTGGAGCGTT GATCCTTACA GCGCCCAAGAA CACATTGTTC TGCACCTGGG ATGAAGTGA GTGCTGTGAG  
  
 SphI  
 ~~~~~  
 A Q G G Q G V N N A H T T Y F G M T S G A C T W
 841 CCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC GCATGCACCT

 NotI
 ~~~~~  
 KDEL  
 ~~~~~  
 . P V A A A E T T E G *
 911 GCGCGGTGCG GCGCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAAA CATTTGGCAA
 981 TAAAGTTTCT TAAGATTGAA TCCTGTGTGCC GGTCTTGCGA TGATTATCAT ATAATTCTG TTGAATTACG
 1051 TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTATTT TATGAGATGG GTTTTATGA TTAGAGTCCC
 1121 GCAATTATAC ATTTAATACG CGATAGAAA CAAATATAG CGCGCAAACT AGGATAAATT ATCGCGCGCG

 HindIII
 ~~~~~  
 ClaI  
 ~~~~~  
 XbaI
 ~~~~~  
 1191 GTGTCACTTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAATTGCG CCTATAGTGA  
 1261 GTCGTATTAC GCGCGCTCAC TGGCCGTGCT TTATACAACGT CCGTACTGGG AAAACCCCTGG CGTTACCCAA  
 1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
 1401 CTTCCCAACA GTTGCAGCAG CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTAA GCGCGGCGGG

FIG. 39C

1471 TGTGGTGGTT ACGGCAGCG TGACCGCTAC ACTTGGCAGC GCCCTAGCG CCGCTCCCTT CGCTTCTCTC  
 1541 CCTTCCCTTC TCGCCACGTT CGCCGGCTTT CCGCGCTAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
 1611 GATTAGTGC TTACGGGCAC CTCGACCCCA AAAAATTGA TTAGGTGAT GGTTCACGTA GTGGCCATC  
 1681 GCCCTGATAG ACGGTTTTC GCCCTTTGAC GCGCTTTGAC GTTGGATCC ACCTTCTTAA ATAGTGGACT CTTGTTCCAA  
 1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATCTCTTTG TTTTCTTTTG ATTATAAGG GATTTTGGCG ATTTCCGGCT  
 1821 ATTGGTTAAA AATGAGCTG AATTAACAA AATTAACGC GAATTTTAA CAAATATTA CCGTTACAAAT  
 1891 TTAGGTGGCA CTTTTCGGG AATGCTGGC GAACCTCTA TTTGTTTATT TTTCTAAATA CATTCAATA  
 1961 TGTATCCGCT CATGAGACAA TAACCTGTAT AATGCTTCA AATGGAAGA AAAGGAAGA GTATGAGTAT  
 2031 TCAACATTTT CGTGTCGCC TTATTTCCCTT TGCTGAAGT TTTTGGCTTC TTACATCGAA CTGCCATCTCA  
 2101 ACGTGGGTGA AAGTAAAGA GATCCTTGAG AGTTTTCGCC CAGTGGGTG TTTTGGCTTC CACGAGTGG  
 2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAAC GGTTCCAATG ATGAGCACTT TTAAGTTCT  
 2241 GCTATGTGGC GCGGTATTAT CCGGTATTGA CCGCGGCA GAGCAACTCG GTCCCGCAT ACCTATTTCT  
 2311 CAGAAATGACT TGGTTGATTA CTCACCAATC ATGAGTGATA ACCTGCGG CCACTTACTT CTGACAACGA TCGGAGGACC  
 2381 TATGCAGTGC TGCATAACC ATGAGTGATA ACCTGCGG CCACTTACTT CTGACAACGA TCGGAGGACC  
 2451 GAAGGAGCTA ACCGCTTTT TGCACCAAT GCGGATCAT GTAACTCGCT TTGATCGTTG GGAACCGGAG  
 2521 CTGAATGAAG CCATACCAAA CGACGAGCGT GACACCAAG TGCCTGTAGC AATGGCAACA AGCTTGGCA  
 2591 AACTATTAACT TGGCGAACTA CTTACTCTAG CTTCCGGCA ACAATTATA GACTGGATGG AGCGGATTA  
 2661 AGTTTCAGGA CCCTTCTGCG GCTCCGGCTT CATTCGAGCA CTGGGGCCAG ATGGTAAGC CTCCCGTATC TGGAGCCGGT  
 2731 GAGGTGGGT CTGCGGGTAT GAGTCAGGCA ACTATGGATG AACGAATAG ACAGATCGCT GAGTAGGTG CCTCACTGAT  
 2801 ACACGACGGG TAAGCATTTG TCTAGGTGAA GATCCCTTTT GATAATCTCA TGACCAAAAT CTTTAGATTG ATTTAAACT GAGTTTCTCT  
 2871 TTTAAAAGGA TCCACTGAGC GTCAGACCC CAAACAAA AACCACCGCT ACCAGCGTG TTTCTGAGAT TCTTAAACT GAGTTTCTCT  
 2941 CTGCTGCTTG CTTTTCCTGA AGGTAACCTG CTTCAGCAGA GCGAGATAC CAAATACTGT CTTCTAGTG TAGCCGTAGT  
 3011 TTTTTCCTGA TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC  
 3081 CTGCTGCTTG CTTTTCCTGA AGGTAACCTG CTTCAGCAGA GCGAGATAC CAAATACTGT CTTCTAGTG TAGCCGTAGT  
 3151 TTTTTCCTGA TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC  
 3221 TGTGCTCAGT GCGGATAGT GCGGATAGT TCTGTCTTAC CCGGTTGGAC TCAAGACGAT TCAAGACGAT TAAGGCGCAG  
 3291 CCGTGGGCT GAACGGGGG TGAGCTATGA GAAAGCGCA CGCTTCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAA  
 3361 ACCTACAGCG GGAACAGGAG AGCGCACGAG GAGCTTCCA GAGGAAACG GAGGAAACG CTTGGTATCT TTATAGTCT  
 3431 CCGCAGGCTC GCGGCTCTG ACCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCTG GAGGAAACG CTTGGTATCT TTATAGTCT  
 3501 GCGGCTCTG GCGGCTCTG ACCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCTG GAGGAAACG CTTGGTATCT TTATAGTCT  
 3571 GCGGCTCTG GCGGCTCTG ACCACCTCTG ACTTGAGCGT CGATTTTGT GATGCTCTG GAGGAAACG CTTGGTATCT TTATAGTCT  
 3641 AAAACGCCAG CAACGCCG CTTTATACGTT TTTTATACGTT TTTTATACGTT TTTTATACGTT TTTTATACGTT TTTTATACGTT  
 3711 TCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC CCGGTTATCC  
 3781 GGAACGACCG AGCGACCG GTCAGTGGC GAGGAGCGG AAGAGCGCC AATACGCAA AATACGCAA AATACGCAA  
 3851 CCGGCTCTG GCGGCTCTG GCGGCTCTG GCGGCTCTG GCGGCTCTG GCGGCTCTG GCGGCTCTG GCGGCTCTG

FIG. 39D

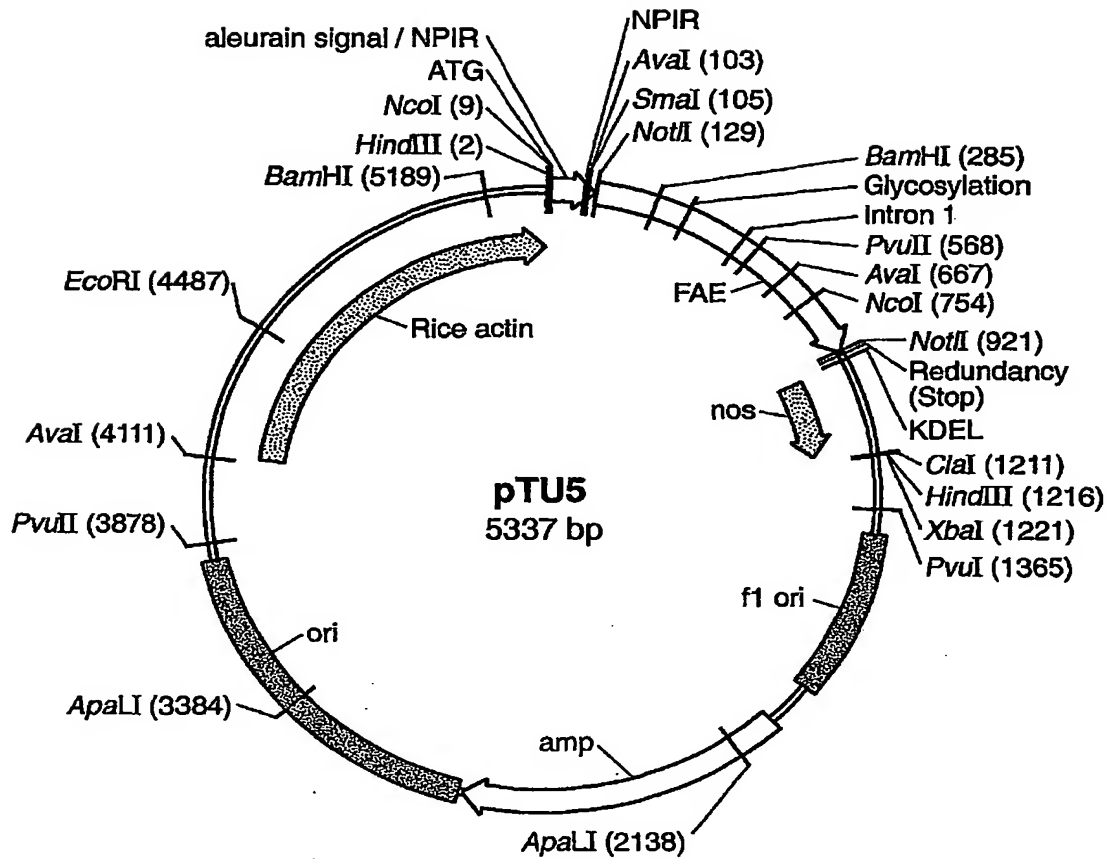


```

3921 CAACGGCAATT AATGTGAGTAT AGTCACTCA TTAGGCACCC CAGGCTTTAC ACTTTATGCT TCCGGCTCGT
3991 ATGTTGTGTG GAATTGTGAG TTTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC
                                     KpnI
                                     XhoI
4061 GCGCAATTAA CCTCACTAA AGGGAACAAA AGCTGGGTAC CGGGCCCCC CTGAGGTCA TTCTATATGCT
4131 TGAGAAGAGA GTCGGATAG TCCAAAATAA AACAAAGTA AGATTACCTG GTCAAAAGTG AAAACATCAG
4201 TTAAAAGGTG GTATAAGTAA AATATCGGTA ATAAAAGGTG GCCCAAAGTG AAATTACTC TTTTCTACTA
4271 TTATAAAAAAT TGAGGATGTT TTGTTCGGTAC TTTGATACGT CATTTTGTGA TGAATGGTT TTTTAAAGTTTA
4341 TTCGCGAATT GGAATGCAT ATCTGTATT GAGTCGGTAT TTAAGTTCGT TGCTTTGTGA AATACAGAGG
4411 GATTGTGATA AGAATATCT TTAAAAAAC CATATGCTAA TTGACATAA TTTTGTGAGAA AAATATATAT
                                     EcoRI
4481 TCAGGCGAAT TCCACATGA ACAATAATA GATTAAAAA GCTTGCCCC GTTGCAGCGA TGGGTATTTT
4551 TTCTAGTAAA ATAAAAGATA AACTTAGACT CAAACATTT ACAAAAACAA CCCCTAAAGT CCTAAGCCC
4621 AAAGTGCTAT GCACGATCCA TAGCAAGCCC AGCCCAACCC AACCCACCCC AACCCACCCC AGTGCAGCCA
4691 ACTGGCAAT AGTCTCCACC CCCGGCACTA TCACCGTGAG TTGTCCGCAC CACCGCACGT CTCGCAGCCA
4761 AAAAAAAA AAGAAAGAAA AAAAAAGAAA AGAAAACAG CAGGTGGGTC CAGGTCTGTG GGGCCGAAA
4831 AGCAGGAGG ATCGGAGCA GCGACGAGC CCGGCCCTCC CTCGGCTTC AAAGAAACGC CCCCCTCGC
4901 CACTATATAC ATACCCCCC CTCTCCTCCC ATCCCCCCTC CCAACACCA ACCACACCA CCACCTCTC
4971 CCCCCTCGCT GCCGACGAC GAGCTCCTCC CCCCCTCCCT TCCGCGCCG CCGGTAAACA CCCCCTCTC
5041 CTCTCTCTTC TTTCCTCGTT TTTTCTTTCG TCTCGGCTC GATCTTTGGC CTGTGTAGTT TGGGTGGGCG
5111 AGAGCGGCTT CGTGGCCAG ATCGGTGCG GGAAGGGGCG GGAATCTGCG GCTGGCGTCT CCGGGCGTGA
                                     BamHI
                                     BglI
5181 GTCGGCCCCG ATCTCGCGG GGAATGGGC TCCTCGGATGT AGATCTTCTT TCCTTCTTCT TTTTGTGGTA
5251 GAATTTGAAT CCTCAGCAT TGTTCATCGG TAGTTTTTCT TTTTCATGAT TGTGACAAAT GCAGCCTCGT
5321 GCGGAGCTTT TTGTAGC

```

FIG. 39E

**FIG.\_40A**

**HindIII NcoI**

~~~~~

1 AAGCTTACCA TGGCCCACGC CCGCGTCCTC CTCCTGGCGC TCGCCGTGCT
TTCGAATGGT ACCGGGTGCG GGCAGGAG GAGGACCGCG AGCGGCACGA

51 GGCCACGGCC GCCGTCGCCG TCGCCTCCTC CTCCTCCTTC GCCGACTCCA
CCGGTGCCGG CCGCAGCGGC AGCGGAGGAG GAGGAGGAAG CCGCTGAGGT

SmaI

~~~~~

**AvaI**

~~~~~

NotI

~~~~~

101 ACCCGGGCCG GCCCGTCACC GACCGCGCGG CCGCCTCCAC GCAGGGCATC  
TGGGCCCCGGC CGGGCAGTGG CTGGCGCGCC GCGGAGGTG CGTCCCGTAG

151 TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC  
AGGCTTCTGG AGATGTCGGC AAATCAGCTT TACCGGTGAT AGAGGGTTCG

201 TGCCTACGCC GACCTGTGCA ACATTCGCTC GACTATTATC AAGGGAGAGA  
ACGGATGCGG CTGGACACGT TGTAAAGCAG CTGATAATAG TTCCCTCTCT

**BamHI**

~~~~~

251 AAATTTACAA TTCTCAAAC T GACATTAACG GATGGATCCT CCGCGACGAC
TTTAAATGTT AAGAGTTTGA CTGTAATTGC CTACCTAGGA GCGGCTGCTG

301 AGCAGCAAAG AAATAATCAC CGTCCTCCGT GGCACCTGGA GTGATACGAA
TCGTCGTTTC TTTATTAGTG GCAGAAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAAC TC GATACTAACT ACACCCTCAC GCCTTTCGAC ACCCTACCAC
AGATGTTGAG CTATGATTGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG

401 AATGCAACGG TTGTGAAGTA CACGGTGGAT ATTATATTGG ATGGGTCTCC
TTACGTTGCC AACACTTCAT GTGCCACCTA TAATATAACC TACCCAGAGG

451 GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA GCCAGTATCC
CAGGTCTTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT CGGTCATAGG

501 GGACTACGCG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCGG
CCTGATGCGC GACTGGCACT GGCCGGTGMG GGAGCCGCGG AGGGACCGCC

PvuII

~~~~~

551 CACTCACTGC CGCCCAGCTG TCTGCGACAT ACGACAACAT CCGCCTGTAC  
GTGAGTGACG GCGGGTCGAC AGACGCTGTA TGCTGTTGTA GCGGACATG

601 ACCTTCGGCG AACC GCGCAG CGGCAATCAG GCCTTCGCGT CGTACATGAA  
TGGAAGCCGC TTGGCGCGTC GCCGTTAGTC CGGAAGCGCA GCATGTACTT

**AvaI**

~~~~~

651 CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA
GCTACGGAAG GTTCGGAGCT CGGGTCCTATG CTGCGTCATA AAGGCCAGT

FIG. 40B

SUBSTITUTE SHEET (RULE 26)

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCCCAGTGGG GCAGGGGTAC
 GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG

 NcoI
 ~~~~~  
 751 GCCCATGGCG GTGTAGAGTA CTGGAGCGTT GATCCTTACA GCGCCCAGAA  
 CGGGTACCGC CACATCTCAT GACCTCGCAA CTAGGAATGT CGCGGGTCTT  
  
 801 CACATTTGTC TGCCTGGGG ATGAAGTGCA GTGCTGTGAG GCCCAGGGCG  
 GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC CGGGTCCCCG  
  
 851 GACAGGGTGT GAATAATGCG CACACGACTT ATTTTGGGAT GACGAGCGGC  
 CTGTCCCACA CTTATTACGC GTGTGCTGAA TAAAACCTA CTGCTCGCCG  
  
 NotI  
 ~~~~~  
 901 GCATGCACCT GGCCGGTTCG GGCCGCGGAA CCACTGAAGG ATGAGCTGTA
 CGTACGTGGA CCGGCCAGCG CCGGCGCCTT GGTGACTTCC TACTCGACAT

 951 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT
 TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA

 1001 CCTGTTGCCG GTCTTGCGAT GATTATCATA TAATTTCTGT TGAATTACGT
 GGACAACGGC CAGAACGCTA CTAATAGTAT ATTAAAGACA ACTTAATGCA

 1051 TAAGCATGTA ATAATTAACA TGTAATGCAT GACGTTATTT ATGAGATGGG
 ATTCGTACAT TATTAATTGT ACATTACGTA CTGCAATAAA TACTCTACCC

 1101 TTTTATGAT TAGAGTCCCG CAATFATACA TTTAATACGC GATAGAAAAC
 AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG

 1151 AAAATATAGC GCGCAAATA GGATAAATTA TCGCGCGCGG TGTATCTAT
 TTTTATATCG CGCGTTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA

 XbaI
 ~~~~~  
 ClaI HindIII  
 ~~~~~  
 1201 GTTACTAGAT CGATAAGCTT CTAGAGCGGC CGGTGGAGCT CCAATTCGCC
 CAATGATCTA GCTATTGAA GATCTCGCCG GCCACCTCGA GGTAAAGCGG

 1251 CTATAGTGAG TCGTATTACG CGCGCTCACT GGCCGTCGTT TTACAACGTC
 GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG

 1301 GTGACTGGGA AAACCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT
 CACTGACCTT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA

 PvuII
 ~~~~~  
 1351 CCCCTTTTCG CCAGCTGGCG TAATAGCGAA GAGGCCCGCA CCGATCGCCC  
 GGGGGAAAGC GGTCGACCGC ATTATCGCTT CTCCGGGCGT GGCTAGCGGG  
  
 1401 TTCCCAACAG TTGCGCAGCC TGAATGGCGA ATGGGACGCG CCCTGTAGCG  
 AAGGGTTGTC AACGCGTCGG ACTTACCGCT TACCCTGCGC GGGACATCGC

**FIG.\_40C**

SUBSTITUTE SHEET (RULE 26)

|                |                           |                          |                          |                          |                           |
|----------------|---------------------------|--------------------------|--------------------------|--------------------------|---------------------------|
| 1451           | GCGCATTAAAG<br>CGCGTAATTC | CGCGGCGGGT<br>GCGCCGCCCA | GTGGTGGTTA<br>CACCACCAAT | CGCGCAGCGT<br>GCGCGTCGCA | GACCGCTACA<br>CTGGCGATGT  |
| 1501           | CTTGCCAGCG<br>GAACGGTCCG  | CCCTAGCGCC<br>GGGATCGCGG | CGCTCCTTTC<br>GCGAGGAAAG | GCTTTCCTTC<br>CGAAAGAAGG | CTTCCTTTCT<br>GAAGGAAAGA  |
| 1551           | CGCCACGTTT<br>GCGGTGCAAG  | GCCGGCTTTC<br>CGGCCGAAAG | CCCGTCAAGC<br>GGGCAGTTCG | TCTAAATCGG<br>AGATTTAGCC | GGGCTCCCTT<br>CCCGAGGGAA  |
| 1601           | TAGGGTTCCG<br>ATCCCAAGGC  | ATTTAGTGCT<br>TAAATCACGA | TTACGGCACC<br>AATGCCGTGG | TCGACCCCAA<br>AGCTGGGGTT | AAAACCTTGAT<br>TTTTGAACFA |
| 1651           | TAGGGTGATG<br>ATCCCACTAC  | GTTACGCTAG<br>CAAGTGCATC | TGGGCCATCG<br>ACCCGGTAGC | CCCTGATAGA<br>GGGACTATCT | CGGTTTTTTCG<br>GCCAAAAAGC |
| 1701           | CCCTTTGACG<br>GGGAAACTGC  | TTGGAGTCCA<br>AACCTCAGGT | CGTCTTTTAA<br>GCAAGAAATT | TAGTGGACTC<br>ATCACCTGAG | TTGTTCCAAA<br>AACAAGGTTT  |
| 1751           | CTGGAACAAC<br>GACCTTGTTG  | ACTCAACCTT<br>TGAGTTGGGA | ATCTCGGTCT<br>TAGAGCCAGA | ATTCTTTTGA<br>TAAGAAACT  | TTTATAAGGG<br>AAATATTCCC  |
| 1801           | ATTTTGCCGA<br>TAAACCGGT   | TTTCGGCCTA<br>AAAGCCGGAT | TTGGTTAAAA<br>AACCAATTTT | AATGAGCTGA<br>TTACTCGACT | TTTAACAAAA<br>AAATTGTTTT  |
| 1851           | ATTTAACGCG<br>TAAATTGCGC  | AATTTTAACA<br>TTAAAATTGT | AAATATTAAC<br>TTTATAATTG | GCTTACAATT<br>CGAATGTTAA | TAGGTGGCAC<br>ATCCACCGTG  |
| 1901           | TTTTTCGGGA<br>AAAAGCCCTT  | AATGTGCGCG<br>TTACACGCGC | GAACCCCTAT<br>CTTGGGGATA | TTGTTTATTT<br>AACAAATAAA | TTCTAAATAC<br>AAGATTTATG  |
| 1951           | ATTCAAATAT<br>TAAGTTTATA  | GTATCCGCTC<br>CATAGGCGAG | ATGAGACAAT<br>TACTCTGTTA | AACCCTGATA<br>TTGGGACTAT | AATGCTTCAA<br>TTACGAAGTT  |
| 2001           | TAATATTGAA<br>ATTATAACTT  | AAAGGAAGAG<br>TTTCCTTCTC | TATGAGTATT<br>ATACTCATAA | CAACATTTCC<br>GTTGTAAAGG | GTGTGCGCCT<br>CACAGCGGGA  |
| 2051           | TATTCCCTTT<br>ATAAGGGAAA  | TTTGCGGCAT<br>AAACGCCGTA | TTTGCCTTCC<br>AAACGGAAGG | TGTTTTTGCT<br>ACAAAAACGA | CACCCAGAAA<br>GTGGGTCTTT  |
| ApaLI<br>~~~~~ |                           |                          |                          |                          |                           |
| 2101           | CGCTGGTGAA<br>GCGACCACTT  | AGTAAAAGAT<br>TCATTTTCTA | GCTGAAGATC<br>CGACTTCTAG | AGTTGGGTGC<br>TCAACCCACG | ACGAGTGGGT<br>TGCTCACCCA  |
| 2151           | TACATCGAAC<br>ATGTAGCTTG  | TGGATCTCAA<br>ACCTAGAGTT | CAGCGGTAAG<br>GTCGCCATTG | ATCCTTGAGA<br>TAGGAACCTT | GTTTTCGCCC<br>CAAAAGCGGG  |
| 2201           | CGAAGAACGT<br>GCTTCTTGCA  | TTTCCAATGA<br>AAAGGTTACT | TGAGCACTTT<br>ACTCGTGAAA | TAAAGTTCTG<br>ATFTCAAGAC | CTATGTGGCG<br>GATACACCGC  |
| 2251           | CGGTATTATC<br>GCCATAATAG  | CCGTATTGAC<br>GGCATAACTG | GCCGGGCAAG<br>CGGCCCGTTC | AGCAACTCGG<br>TCGTTGAGCC | TCGCCGCATA<br>AGCGGCGTAT  |
| 2301           | CACTATTCTC<br>GTGATAAGAG  | AGAATGACTT<br>TCTTACTGAA | GGTTGAGTAC<br>CCAACTCATG | TCACCAGTCA<br>AGTGGTCAGT | CAGAAAAGCA<br>GTCTTTTCGT  |

**FIG. 40D**

SUBSTITUTE SHEET (RULE 26)

2351 TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA  
AGAATGCCTA CCGTACTGTC ATTCTCTTAA TACGTCACGA CGGTATTGGT

2401 TGAGTGATAA CACTGCGGCC AACTTACTTC TGACAACGAT CGGAGGACCG  
ACTCACTATT GTGACGCCGG TTGAATGAAG ACTGTTGCTA GCCTCCTGGC

2451 AAGGAGCTAA CCGCTTTTTT GCACAACATG GGGGATCATG TAACTCGCCT  
TTCCTCGATT GGCMAAAAAA CGTGTTGTAC CCCCTAGTAC ATTGAGCGGA

2501 TGATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG  
ACTAGCAACC CTTGGCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC

2551 ACACCACGAT GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAAC  
TGTGGTGCTA CGGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA

2601 GCGAACTAC TTA CTCTAGC TTCCCGGCAA CAATTAATAG ACTGGATGGA  
CCGCTTGATG AATGAGATCG AAGGGCCGTT GTTAATTATC TGACCTACCT

2651 GCGGATATAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT  
CCGCTATTT CAACGTCCTG GTGAAGACGC GAGCCGGGAA GGCCGACCGA

2701 GGTTTATTCG TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC  
CCAAATAACG ACTATTTAGA CCTCGGCCAC TCGCACCCAG AGCGCCATAG

2751 ATTGCAGCAC TGGGGCCAGA TGGTAAGCCC TCCCGTATCG TAGTTATCTA  
TAACGTCGTG ACCCCGGTCT ACCATTGCGG AGGGCATAGC ATCAATAGAT

2801 CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA CAGATCGCTG  
GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCTTTATCT GTCTAGCGAC

2851 AGATAGGTGC CTCACTGATT AAGCATTTGGT AACTGTCAGA CCAAGTTTAC  
TCTATCCACG GAGTGACTAA TTCGTAACCA TTGACAGTCT GGTTCAAATG

2901 TCATATATAC TTTAGATTGA TTTAAACTT CATTTTAAAT TTAAAGGAT  
AGTATATATG AAATCTAACT AAATTTTGAA GTAAAAATTA AATTTTCTA

2951 CTAGGTGAAG ATCCTTTTTG ATAATCTCAT GACCAAAATC CCTTAACGTG  
GATCCACTTC TAGGAAAAAC TATTAGAGTA CTGGTTTTAG GGAATTGCAC

3001 AGTTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
TCAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTTCTA GTTTCCTAGA

3051 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAA  
AGAACTCTAG GAAAAAAGA CGCGCATTAG ACGACGAACG TTTGTTTTTT

3101 ACCACCGCTA CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC  
TGGTGGCGAT GGTGCGCCACC AAACAAACGG CCTAGTTCTC GATGGTTGAG

3151 TTTTTCGAA GGTAACGGC TTCAGCAGAG CGCAGATACC AAATACTGTC  
AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG TTTATGACAG

3201 CTTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAACT CTGTAGCACC  
GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTTC TTGA GACATCGTGG

3251 GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG  
CGGATGTATG GAGCGAGACG ATTAGGACAA TGGTCACCGA CGACGGTCAC

**FIG. 40E**

SUBSTITUTE SHEET (RULE 26)

3301 GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT  
CGCTATTTCAG CACAGAATGG CCCAACCTGA GTTCTGCTAT CAATGGCCTA

## ApaLI

~~~~~

3351 AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGCACAC AGCCCAGCTT
TTCCGCGTTCG CCAGCCCAGC TTGCCCCCCA AGCACGTGTG TCGGGTCGAA

3401 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG
CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGATGTCGCA CTCGATACTC

3451 AAAGCGCCAC GCTTCCCGAA GGGAGAAAGG CGGACAGGTA TCCGGTAAGC
TTTCGCGGTG CGAAGGGCTT CCCTCTTTCC GCCTGTCCAT AGGCCATTCG

3501 GGCAGGGTCG GAACAGGAGA GCGCACGAGG GAGCTTCCAG GGGGAAACGC
CCGTCCCAGC CTTGTCTCTT CCGTGCTCC CTCGAAGGTC CCCCCTTTGCG

3551 CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA CTTGAGCGTC
GACCATAGAA ATATCAGGAC AGCCCAAAGC GGTGGAGACT GAACTCGCAG

3601 GATTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC
CTAAAAACAC TACGAGCAGT CCCCCGCCT CGGATACCTT TTTGCGGTGCG

3651 AACGCGGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT
TTGCGCCGGA AAAATGCCAA GGACCGGAA ACGACCGGA AACGAGTGT

3701 GTTCTTTTCT GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT
CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA

3751 TTGAGTGAGC TGATACCGCT CGCCGCAGCC GAACGACCGA GCGCAGCGAG
AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTGCTGGCT CGCGTCGCTC

3801 TCAGTGAGCG AGGAAGCGGA AGAGCGCCCA ATACGCAAAC CGCCTCTCCC
AGTCACTCGC TCCTTCGCCT TCTCGCGGGT TATGCGTTTG GCGGAGAGGG

PvuII

~~~~~

3851 CGCGCGTTGG CCGATTTCATT AATGCAGCTG GCACGACAGG TTTCCCGACT  
GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGCTGTCC AAAGGGCTGA

3901 GGAAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCACTCAT  
CCTTTCGCCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA

3951 TAGGCACCCC AGGCTTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGG  
ATCCGTGGGG TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC

4001 AATTGTGAGC GGATAACAAT TTCACACAGG AAACAGCTAT GACCATGATT  
TTAACACTCG CCTATTGTTA AAGTGTGTCC TTTGTCGATA CTGGTACTAA

4051 ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGGAAACAAA GCTGGGTACC  
TGCGGTTCGC GCGTTAATTG GGAGTGATT CCCTTGTTTT CGACCCATGG

## AvaI

~~~~~

4101 GGGCCCCCCC TCGAGGTCAT TCATATGCTT GAGAAGAGAG TCGGGATAGT
CCCGGGGGGG AGCTCCAGTA AGTATACGAA CTCTTCTCTC AGCCCTATCA

FIG. 40F

SUBSTITUTE SHEET (RULE 26)

```

4151  CCAAAATAAA  ACAAAAGGTAA  GATTACCTGG  TCAAAAAGTGA  AAACATCAGT
      GGTTTTATTT  TGTTCCTCATT  CTAATGGACC  AGTTTTCACT  TTTGTAGTCA

4201  TAAAAGGTGG  TATAAGTAAA  ATATCGGTAA  TAAAAGGTGG  CCCAAAGTGA
      ATTTTCACC  ATATTCATTT  TATAGCCATT  ATTTTCACC  GGGTTTCACT

4251  AATTTACTCT  TTTCTACTAT  TATAAAAATT  GAGGATGTTT  TGTCCGTACT
      TTAAATGAGA  AAAGATGATA  ATATTTTAA  CTCCTACAAA  ACAGCCATGA

4301  TTGATACGTC  ATTTTGTAT  GAATTGGTTT  TTAAGTTTAT  TCGCGATTTG
      AACTATGCAG  TAAAAACATA  CTTAACCAAA  AATTCAAATA  AGCGCTAAAC

4351  GAAATGCATA  TCTGTATTTG  AGTCGGTTTT  TAAGTTCGTT  GCTTTTGTAA
      CTTTACGTAT  AGACATAAAC  TCAGCCAAAA  ATTCAAGCAA  CGAAAACATT

4401  ATACAGAGGG  ATTTGTATAA  GAAATATCTT  TAAAAAACCC  ATATGCTAAT
      TATGTCCTCC  TAAACATATT  CTTTATAGAA  ATTTTTTGGG  TATACGATTA

                                     EcoRI
                                     ~~~~~~

4451  TTGACATAAT  TTTTGAGAAA  AATATATATT  CAGGCGAATT  CCACAATGAA
      AACTGTATTA  AAAACTCTTT  TTATATATAA  GTCCGCTTAA  GGTGTTACTT

4501  CAATAATAAG  ATTAATAATAG  CTTGCCCCCG  TTGCAGCGAT  GGGTATTTT
      GTTATTATTC  TAATTTTATC  GAACGGGGGC  AACGTCGCTA  CCCATAAAAA

4551  TCTAGTAAAA  TAAAAGATAA  ACTTAGACTC  AAAACATTTA  CAAAAACAAC
      AGATCATTTT  ATTTTCTATT  TGAATCTGAG  TTTTGTAAAT  GTTTTGTGTTG

4601  CCCTAAAGTC  CTAAAGCCCA  AAGTGCTATG  CACGATCCAT  AGCAAGCCCA
      GGGATFTTCAG  GATTTCTGGGT  TTCACGATAC  GTGCTAGGTA  TCGTTCGGGT

4651  GCCCAACCCA  ACCCAACCCA  ACCCACCCA  GTGCAGCCAA  CTGGCAAATA
      CGGGTTGGGT  TGGGTGGGT  TGGGTGGGT  CACGTCGGT  GACCGTTTAT

4701  GTCTCCACCC  CCGGCACTAT  CACCGTGAGT  TGTCCGCACC  ACCGCACGTC
      CAGAGGTGGG  GGCCGTGATA  GTGGCACTCA  ACAGGCGTGG  TGGCGTGCAG

4751  TCGCAGCCAA  AAAAAAAAAA  AGAAAGAAAA  AAAAGAAAA  GAAAAACAGC
      AGCGTCGGTT  TTTTTTTTTT  TCTTCTTTT  TTTTCTTTT  CTTTTGTGTCG

4801  AGGTGGGTCC  GGGTCGTGGG  GGCCGGAAAA  GCGAGGAGGA  TCGCGAGCAG
      TCCACCCAGG  CCCAGCACCC  CCGGCCTTTT  CGCTCCTCCT  AGCGCTCGTC

4851  CGACGAGGCC  CGGCCCTCCC  TCCGCTTCCA  AAGAAACGCC  CCCCATCGCC
      GCTGCTCCGG  GCCGGGAGGG  AGGCGAAGGT  TTCCTTTCGG  GGGGTAGCGG

4901  ACTATATACA  TACCCCCCCC  TCTCCTCCCA  TCCCCCAAC  CCTACCACCA
      TGATATATGT  ATGGGGGGGG  AGAGGAGGGT  AGGGGGGTTG  GGATGGTGGT

4951  CCACCACCAC  CACCTCCTCC  CCCCTCGCTG  CCGGACGACG  AGCTCCTCCC
      GGTGGTGGTG  GTGGAGGAGG  GGGGAGCGAC  GGCTGCTGC  TCGAGGAGGG

5001  CCCTCCCCCT  CCGCGCGCGC  CGGTAACCAC  CCCGCCCTC  TCCTCTTTCT
      GGGAGGGGGA  GGCAGCGCGC  GCCATTGGTG  GGGCGGGGAG  AGGAGAAAGA

```

FIG. 40G

SUBSTITUTE SHEET (RULE 26)

5051 TTCTCCGTTT TTTTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT
AAGAGGCAAA AAAAAAGCA GAGCCAGAGC TAGAAACCGG AACCATCAAA

5101 GGGTGGGCGA GAGCGGCTTC GTCGCCAGA TCGGTGCGCG GGAGGGGCGG
CCCACCCGCT CTCGCCGAAG CAGCGGGTCT AGCCACGCGC CCTCCCCGCC

BamHI
~~~~~

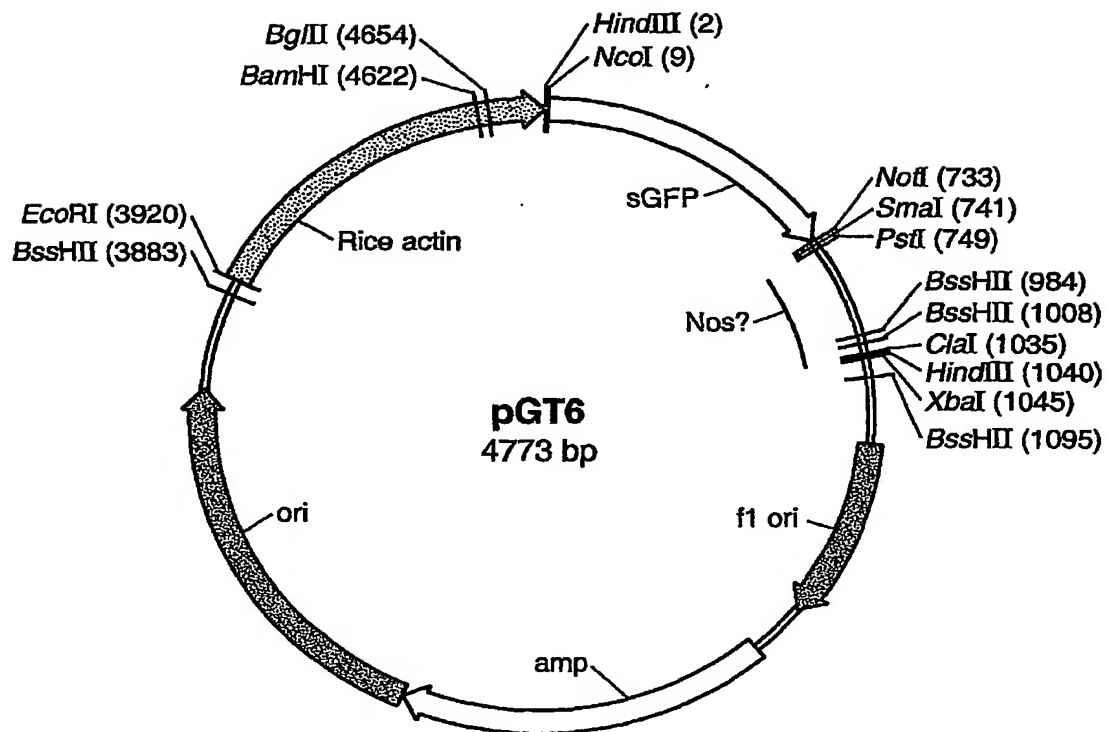
5151 GATCTCGCGG CTGGCGTCTC CGGGCGTGAG TCGGCCCGGA TCCTCGCGGG  
CTAGAGCGCC GACCGCAGAG GCCCGCACTC AGCCGGGCCT AGGAGCGCCC

5201 GAATGGGGCT CTCGGATGTA GATCTTCTTT CTTTCTTCTT TTTGTGGTAG  
CTTACCCCGA GAGCCTACAT CTAGAAGAAA GAAAGAAGAA AAACACCATC

5251 AATTGAATC CCTCAGCATT GTTCATCGGT AGTTTTCCTT TTCATGATTT  
TTAAACTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAGAA AAGTACTAAA

5301 GTGACAAATG CAGCCTCGTG CGGAGCTTTT TTGTAGC  
CACTGTTTAC GTCGGAGCAC GCCTCGAAAA AACATCG

**FIG.\_40H**

**FIG.\_41A**

```

HindIII NcoI
1  AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCCG GGGTGGTGCC CATCTGGTC GAGCTGGACG
   TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAAGTGGC CCCACACAGG GTAGGACCAG CTCGACCTGC

71  GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT GCCACCTACG GCAAGCTGAC
   CGCTGCACCT GCCGGTGTTC AAGTCGCACA GGGCGCTCCC GCTCCCGCTA CGGTGGATGC CGTTGCACTG

141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCACCCC TCGTGACCAC CTTACACCTAC
   GGACTTCAAG TAGACGTGGT GGCCGTTTGA CCGGCACGGG ACCGGGTGGG AGCACTGGTG GAAGTGGATG

211 GGGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT CTTCAAGTCC GCCATGCCCG
   CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGTGAA GAAATTGAGG CGGTACGGGC

281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CCGCAACTAC AAGACCCGCG CCGAGGTGAA
   TTCCGATGCA GGTCTTCGGG TGGTAGAAGA AGTTCCCTGCT GCCGTTGATG TTCTGGGCGC GGCTCCACTT

351 GTTCGAGGGC GACACCTTGG TGAACCGCAT CGAGTGAAG GGCATCGACT TCAAGGAGGA CGGCAACATC
   CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA AGTTCTCCTT GCCGTTGTAG

421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT GGCCTGACAAG CAGAAGAACG
   GACCCCGTGT TCGACCTCAT GTTGATGTGG TCGGTGTTGC AGATATAGTA CCGGCTGTTC GTCTTCTTGC

491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CCGCAGCGTG CAGCTGCGCG ACCACTACCA
   CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTCCAC GTGAGCGCGC TGGTGATGGT

561 GCAGAACACC CCCATCGGGC ACGGCCCGGT GCTGCTGCCC GACACCACT ACCTGAGCAC CCAGTCCGCC
   CGTCTTGTGG GGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGTTGA TGGACTCGTG GGTGAGGCGG

631 CTGAGCAAAG ACCCCACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT CGTGACCGCC GCCGGGATCA
   GACTCGTTTC TGGGGTTGCT CTTGCGCGTA GTGTACCAGG ACGACCTCAA GCACTGGCGG CGGCCCTAGT

```

**FIG. 41B**

```

SmaI
~~~~~
NotI PstI
~~~~~
701 CTCACGGCAT GGACGAGCTG TACAAGTAA CGGCGCGCCC GGGCTGCAGG GAAACCAC TG AAGGATGAGC
    GAGTGCCTGA CCTGCTCGAC ATGTTCAATT CGCGCGCGGG CCGACGCTCC CTTTGCTGAC TTCCTACTCG

771 TGTAAAGRAG CAGATCGTTC AAACATTTGG CAATAAGTT TCTTAAGATT GAATCCTGTT GCCGGTCTTG
    ACATTTCTTC GTCTAGCAAG TTTGTAAACC GTTATTTCAA AGAATCTAA CTTAGGACAA CGGCCAGAAC

841 CGATGATTAT CATATAATT CTGTTGAATT ACGTTAAGCA TGTAAATAAT AACATGTAAT GCATGACGTT
    GCTACTAATA GTATATTAAA GACAACTTAA TGCAATTCGT ACATATTAA TTGTACATTA CGTACTGCAA

911 ATTTATGAGA TGGGTTTTTA TGATTAGAGT CCCGCAATTA TACATTTAAT ACGCGATAGA AAACAAAAATA
    TAAATACTCT ACCAAAAAT ACTAATCTCA GCGCGTTAAT ATGTAATAA TCGCGTATCT TTTGTTTTAT

                                ClaI HindIII
                                ~~~~~
 XbaI
                                ~~~~~
                                BssHII
                                ~~~~~
981 TAGCGCGCAA ACTAGGATAA ATTATCGCG CGGTGTCAT CTATGTTACT AGATCGATA GCTTCTAGAG
 ATCGCGCGTT TGATCCTATT TAATAGCGCG CGCCACAGTA GATACAATGA TCTAGCTATT CGAAGATCTC

 BssHII
                                ~~~~~
1051 CGGCCGGTGG AGCTCCAATT CGCCCTATAG TGAGTCGTAT TACGCGCGCT CACTGGCCGT CGTTTACAA
    GCCGGCCACC TCGAGGTTAA GCGGATATC ACTCAGCATA ATGCGCGCGA GTGACCGGCA GCATAATGTT

1121 CGTCGTGACT GGGAAAACCC TGCGGTTACC CAACTTAATC GCCTTGCAGC ACATCCCCCT TTCGCCAGCT
    GCAGCACTGA CCTTTTGGG ACCGCAATGG GTTGAATTAG CGGAACGTCG TGTAGGGGGA AAGCGGTCCA

1191 GCGGTAATAG CGAAGAGGCC CGCACCGATC GCCCTTCCCA ACACTTGC GCCTGAATG GCGAATGGGA
    CCGCATTATC GCTTCTCCGG GCGTGGCTAG CCGGAAGGGT TGTCAACGCG TCGGACTTAC CGCTTACCCCT

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA GCGTGACCGC TACACTTGCC
    GCGCGGGACA TCGCCGCGTA ATTGCGCGCG CCCACACCAC CAATGCGCGT CGCACTGGCG ATGTGAACGG

```

FIG. 41C

1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC GTTCGCCGGC TTTCCCCCGTC  
TCGCGGGATC GCGGGCGAGG AAGCGAAG AAGGCGGTG AAGAGCGGTG CAAGCGGCGG AAAGGGGCAG  
1401 AAGCTCTAAA TCGGGGCTC CCTTTAGGGT TCCGATTTAG TCGTTTACGG CACCTCGACC CCAAAAAACT  
TTCGAGATTT AGCCCCGAG GGAATCCCA AGCTAAATC ACGAATGCC GTGGAGCTGG GGTTTTTTGA  
1471 TGATTAGGGT GATGTTTAC GTAGTGGCC ATCGCCCTGA TAGACGGTTT TTGCCCCTTT GACGTTGGAG  
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGACT ATCGCCAAA AAGCGGAAA CTGCAACCTC  
1541 TCCACGTTCT TTAATAGTGG ACTCTTGTTC CAAACTGGAA CAACACTCAA CCTATCTCG GTCTATTTCTT  
AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT GGGATAGAGC CAGATAAGAA  
1611 TTGATTATA AGGATTTTG CCGATTTCCG CCTATTGGTT AAAAAATGAG CTGATTTAAC AAAAATTAA  
AACTAAATAT TCCCTAAAAC GGCTAAAGCC GGATAACCAA TTTTCTACTC GACTAAATTG TTTTAAATTT  
1681 CGCGAATTTT AACAAATAT TAACGCTTAC AATTTAGGTG GCCTTTTCG GGGAAATGNG CGCGGAACCC  
GCGCTTAAA TTGTTTATA ATTCGGAATG TTAATCCAC CGTGAAGAAG CCCTTTTACAC GCGCCTTGGG  
1751 CTATTGTGTT ATTTTCTTAA ATACATTCAA ATATGTATCC GCTCATGAGA CAATAACCTT GATAAATGCT  
GATAAACAAA TAAAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT GTTATTGGGA CTATTTACGA  
1821 TCAATAATAT TGA AAAAGGA AGATATGAG TATTCAACAT TTCCGTGTCTG CCTTATTCC CTTTTTTGCG  
AGTTATTATA ACTTTTTCCT TCTCATCTC ATAAGTTGTA AAGGCACAGC GGAATAAGG GAAAAACGC  
1891 GCATTTGCCC TTCTGTGTTT TGCTCACCCA GAAACGCTGG TGAAGTAAA AGATGCTGAA GATCAGTTGG  
CGTAAAACGG AAGACAAA ACGAGTGGT CTTTGGGACC ACTTTCAATTT TCTACGACTT CTAGTCAACC  
1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT GAGAGTTTTC GCCCCGAAGA  
CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTCGCC ATTCTAGGAA CTCTCAAAAG CCGGGCTTCT  
2031 ACGTTTCCA ATGATGAGCA CTTTAAAGT TCTGCTATGT GCGCGGTAT TATCCCGTAT TGACGCCGGG  
TGCAAAAGGT TACTACTCGT GAAATTTCA AGACGATACA CCGCGCCATA ATAGGGCATA ACTGCGGCC  
2101 CAAGAGCAAC TCGGTGCGCG CATACACTAT TCTCAGATG ACTTGGTTGA GTACTCACA GTACAGAAA  
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT CATGAGTGGT CAGTGTCTTT

**FIG. 41D**

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG ATAACACTGC  
TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC ACGACGGTAT TGGTACTCAC TATTGTGACG

2241 GGCCAACTTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT TTTTGCACAA CATGGGGGAT  
CCGGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGCGAA AAAACGTGTT GTACCCCTTA

2311 CATGTAACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC AAACGACGAG CGTGACACCA  
GTACATTGAG CGGAAC TAGC AACCTTGGC CTCGACTTAC TTCCGTTATGG TTTGCTGCTC GCACTGTGTT

2381 CGATGCCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGGCGAA CTACTTACTC TAGCTTCCC  
GCTACGGACA TCGTTACCCT TGTTCGAACG CGTTTGATAA TTGACCCGCTT GATGAATGAG ATCGAAGGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGGCTCGGC CCTTCCGGCT  
CGTTGTTAAT TATCTGACCT ACCTCCGCTT ATTTCAACGT CCTGTTGAG ACCTGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCG TATCATTTGCA GCACCTGGGC  
CCGACCAAT AACGACTATT TAGACTCTGG CCACTCGCAC CCAGAGCGCC ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG ATGAACGAAA  
GTCTACCAAT CGGGAGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC CGTTGATACC TACTTGCITT

2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACGT CAGACCAAGT TTACTCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA GTCTGGTTCA AATGAGTATA

2731 ATACTTTAGA TTGATTTAAA ACTTCATTCT TAATTTAAA GGATCTAGGT GAAGATCCTT TTTGATAATC  
TATGAATCT AACTAAATTT TGAAGTAAAA ATTAAATTTT CCTAGATCCA CTTCTAGGAA AAACATATTAG

2801 TCATGACCAA AATCCCTTAA CGTGAGTTT CATTCCACTG AGCTCAGAC CCCGTAGAAA AGATCAAAGG  
AGTACTGGTT TTAGGGAATT GCACTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCITTTT TTCTGGCGGT AATCTGCTGC TTGCAACAA AAAAACACC GCTACCAGCG  
TAGAAGAAT CTAGGAAAAA AAGACGCGCA TTAGACGACG AACGTTTGT TTTTGTGGT CGATGGTCCG

2941 GTGGTTTGTG TGCCGGATCA AGAGCTACCA ACTCTTTTTC CGAAGTTAAC TGGCTTCAGC AGAGCGCAGA  
CACCAAAACA ACGGCCCTAGT TCTCGATGTT TGAGAAAAAG GCTTCCATTG ACCGAAGTCG TCTCGCGTCT

**FIG. 41E**

3011 TACCAAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTTCAAG AACTCTGTAG CACCGCCTAC  
ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG  
3081 ATACCTCGCT CTGCTAATCC TGTACCAGT GGCTGCTGCC AGTGGCGATA AGTCGTGTCT TACCGGGTTG  
TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCGCTAT TCAGCACAGA ATGCCCCAAC  
3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGGTCCG GGTGTCGTGC ACACAGCCCA  
CTGAGTTCTG CTATCAATGG CCTATTCCGC GTCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTGGGT  
3221 GCTTGGAGCG AACGACCTAC ACCGAACCTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC  
CGAACCTCGC TTGCTGGATG TGGCTTGAAT TATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG  
3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGAGCTT  
GCTTCCCTCT TTCCGCCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGTG CTCTCGCGTG CTCCTCGAA  
3361 CCAGGGGGA AGCCTTGGTA TCTTTATAGT CCTGTCCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT  
GGTCCCCCTT TCGGGACCAT AGAAATATCA GGACAGCCCA AAGCGGTGGA GACTGAATC GCAGCTAAAA  
3431 TGTGATGCTC GTCAGGGGG GAGAGCCTAT GGAATAACGC CAGCAACGCG GCCTTTTATC GGTTCCTGGC  
ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG GTCTGTGCG CGGAAAAATG CCNAGGACCG  
3501 CTTTGTCTGG CCTTTTGTCT ACATGTTCTT TCCTGCGTTA TCCCCTGAT CTGTGGATAA CCGTATTACC  
GAAAACGACC GGAACGAG TGTACAAGAA AGGACGCAAT AGGGGACTAA GACACCTATT GGCATAATGG  
3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG  
CGGAAACTCA CTCGACTATG GCGAGCGGC TCGGCTTGCT GGTCTCGCTC GCTCAGTCAC TCGCTCCTTC  
3641 CGGAAGAGCG CCCATACGC AAACCGCTC TCCCCGCGG TTGGCCGATT CATTAATGCA GCTGGCACGA  
GCCTTCTCGC GGGTTATGCG TTGCGCGAG AGGGCGCGC AACCGGCTAA GTAAATTACGT CGACCGTGCT  
3711 CAGGTTTCCC GACTGGAAAG CCGGCAGTGA GCGCAACGCA ATTAATGTA GTTAGCTCAC TCATTAGGCA  
GTCCAAAGG CTGACCTTTC GCCGTCATC CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAATCCGT  
3781 CCCCAGGCTT TACACTTAT GCTTCCGGCT CGTATGTTGT GTGAATTGT GAGCGGATAA CAATTTCACA  
GGGTCCGAA ATGTGAATA CGAAGCCGA GCATACAACA CACCTTAACA CTCGCCTATT GTTAAAGTGT

**FIG. 41F**

```

      BssHII
      ~~~~~
3851 CAGGAACACAG CTATGACCAT GATTACGCCA AGCGCGCAAT TAACCCCTCAC TAAAGGGAAC AAAAGCTGGA
 EcorI
      ~~~~~
      ~~~~~
3921 ATTCCACAAT GAACAATAAT AAGATTAAAA TAGCTTGCCC CCGTTGCAGC GATGGGTATT TTTTCTAGTA
 TAAGGTGTTA CTTGTTATTA TTCTAATTTT ATCGAACGGG GGCAACGTCG CTACCCATAA AAAAGATCAT

3991 AAATAAAGA TAAACTTAGA CTCAAAACAT TTACAAAAC AACCCCTAAA GTCTAAAGC CCAAAAGTGT
 TTTATTTTCT ATTTGNACTT GAGTTTGTGA AATGTTTGT TTGGGGATTT CAGGATTCG GGTTCACGA

4061 ATGCACGATC CATAGCAAGC CCAAGCCAAAC CCAACCCACC CCAAGTCAGC CCAACTGGCAA
 TACGTGCTAG GTATCGTTTC GGTGCGGTTC GGTGCGGTTC GGTACGTCG GTTGACCGTT

4131 ATAGTCTCCA CCCCGGCAC TATCACGCTG AGTTGTCCGC ACCACCGCAC GTCGCGCAGC CAAAAA
 TATCAGAGGT GGGGGCCGTG ATAGTGGCAC TCAACAGGCG TGGTGGCGTG CAGAGCGTCG GTTTT
 TTTTCTTCT TTTTCTTCTT TTTCTTTTG TCGTCCACCC AGGCCCAGCA CCCCAGCCT TTTTCGCTCCT

4201 AAAAGAAAGA AAAAAGAA AAAAAGAA AGCAGGTGG TCCGGGTCTG GGGGGCCGGA AAAGCGAGGA
 TTTTCTTCT TTTTCTTCTT TTTCTTTTG TCGTCCACCC AGGCCCAGCA CCCCAGCCT TTTTCGCTCCT

4271 GGATCGCGAG CAGCGACGAG GCCCGGCCCT CCTCCGCTT CCAAGAAAC GCTCCCTCATC GCCACTATAT
 CCTAGCGCTC GTCGCTGCTC CCGCGCGGGA GGGAGGCGAA GGTTCCTTTG CCGGGGTAG CGGTGATATA

4341 ACATACCCCC CCTCTCTCTC CCATCCCTCC AACCTTACCA CCACCACTCC CACCCCTCTCG
 TGTATGGGGG GGGAGAGGAG GGTAGGGGGG TTGGGATGGT GGTGTTGTTG GTGTTGAGG AGGGGGAGC

4411 CTGCCGGACG ACGAGTCTCT CCCCCTCTCC CCTCCGCGCG CGCCGGTTAC CACCCCGCCC CTCTCTCTCT
 GACGGCTTGC TGCTCGAGGA GGGGGGAGGG GAGGCGGCG GCGGCCATTG GTGGGGCGGG GAGAGGAGA

4481 TCTTCTCTCG TTTTCTTTT TCTCTCGGTC TCGATCTTTG GCCTTGGTAG TTGGGTGGG CGAGAGCGGC
 AGAAAGAGGC AAAAAAAGAA GCAGAGCCAG AGCTAGAAAC CGGAACCATC AAACCAACCC GCTCTCGCCG

4551 TTCTGTCGCC AGATCGGTGC GCGGAGGGG CGGGATCTCG CGGTGGCGT CTCCGGGCGC GAGTCGGGCC
 AAGCAGCGGG TCTAGCCACG CGCCCTCCCC GCCCTAGAGC GCGGACCGCA GAGGCCCGCA CTCAGCCCGG

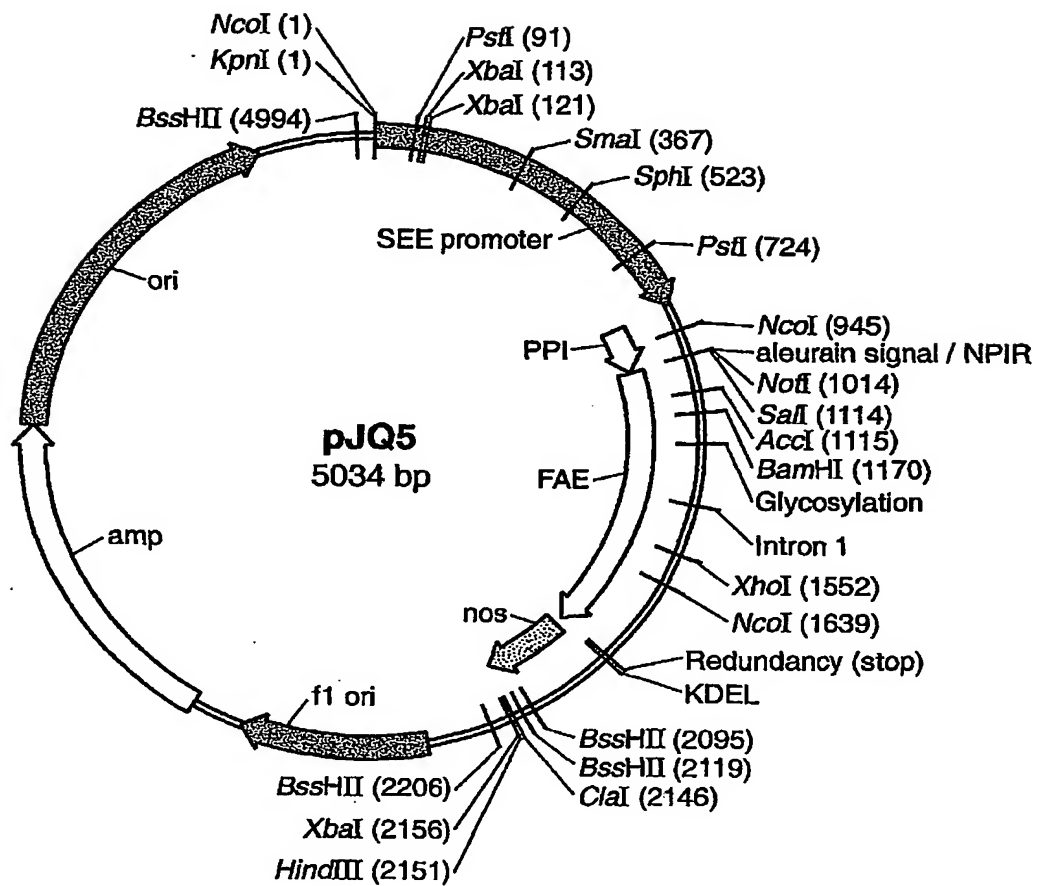
```

FIG. 41G



|      | BamHI                                                                                                                                                         | BglII |
|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
|      | ~~~~~                                                                                                                                                         | ~~~~~ |
| 4621 | GGATCCTCGC GGGGAATGGG GCTCTCGGAT GTAGATCTTC TTTCTTTCTT CTTTTGTGG TAGAATTGA<br>CCTAGGAGCG CCCCTTACCC CGAGAGCCTA CATCTAGAAG AAAGAAAGAA GAAAAACACC ATCTTAAACT    |       |
| 4691 | ATCCCTCAGC ATTGTTTCATC GGTAGTTTTT CTTTTCATGA TTTGTGACAA ATGCAGCCTC GTGCGGAGCT<br>TAGGGAGTCG TAACAAGTAG CCATCAAAA GAAAAGTACT AAACACTGTT TACGTCGGAG CACGCCCTCGA |       |
| 4761 | TTTTGTAGG TAG<br>AAAAACATCC ATC                                                                                                                               |       |

**FIG.\_41H**

**FIG. 42A**

NcoI  
~~~~~  
KpnI  
~

1 CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG  
GTACCCGGTC CATATTAAATA CCTATATAG TCGTTTATT AGCTTTATAG TGTTAACCGA TGTATATAGAC

~~~~~  
PstI XbaI XbaI  
~~~~~

71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAACT CTAGATAGCA CAGCCACAGC  
TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAACA TAGATCTTGA GATCTATCGT GTCGGTGTCTG

141 ACCTACAGGA GTCCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT  
TGGATGTCTT CACGCTGTGA ACACCTGACA TCATCACAAC CTCTGCCCTCG AGAAAGGATG GAGGACTGCA

211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGGGCTC CCAACAAAAT ATCGTCCCCC  
ACGGCGGCAA CAGGTAAGGT TGCCGTATG AGAGTTGGTT AGTGGCGGAG GGTTCGTTTA TAGCAGGGGG

281 ATGTCCTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTGCT GAATCTCGCT TCCACTGGCC  
TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACACGA CTTAGAGCGA AGGTGACCGG

~~~~~  
SmaI  
~~~~~

351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCTG TCACCCCTGG CGTCATGGBA  
TTAGTTCGAGT CGAGGGCCCT CGAGTGAGTA AGTCTAGGG TAGCAGCAGC AGTGGGGACC GCAGTACCCCT

421 TGGAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT  
ACCTTTTCTT GGAGGCAACG AGCCTACTCA GTCCGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA

~~~~~  
SphI  
~~~~~

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATGAC TTTATTTTGG  
AGCTAAAGG GGTATCTCTT TTCAATATCGT ACGAAAGCCC AAAACAAACC GAATTAACAG AAATAAANAAC

561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAF CGAGACGGAT AATAGGCTGG  
AACCTCAACT TAGGACTAAA CAACACATTT TACGGGTTGG TAGACTTATA GCTCTGCCCTA TTATCCGACC

FIG. 42B

```

631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTTACA GCTGGAGGCT
 GATTAAATTAA ATATCGTTCT AGACATCAC GTGTAGCGTT TATAGAAAGA CCCGTAATGT CGACCTCCGA

 PstI
          ~~~~~
701  TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
    AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTCTGTTACC ACTTCGCACC GCTACTCTAC CCATATTTTG

771  CCCCAGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTGCCCCGCT TCCCCCTGCC GGACGACCCA
    GGGCCCGTGG CCTTGCCTC GAGGCGGAT GGTCAATGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841  GTAAAATACT GTTGCCCACT CGCCGCGGAG ATGGMCGTGC ACAAGGAGGT SAATCTCGTS GCCTACCTCC
    CATTTATGA CAACGGGTGA GCGGCCGCTC TACCKGCACG TGTTCCTCCA TTGAAGCAS CGGATGGAGG

          NcoI
          ~~~~~
911 TGATCGTCTC CGGCCCTCCTC TTGCTCGTST CCGCCATGGA GCACGTGGAG GCCAAGGCTT GCACCCKCGA
 ACTAGCAGGA GCCGGAGGAG AACGAGCASA GCGGCTACCT CGTGCACCTG CGGTTCCGGA CGTGGGCGCT

 NotI
          ~~~~~
981  GTGCGGCAAC CTCGGCTTCG GCATCTGCCC GCGGCGCCGC TCCACGCGAG GCATCTCCGA AGACCTCTAC
    CACGCCGTTG GAGCCGAAGC CGTAGACGGG CCGCCGCGCG AGGTGCGTCC CGTAGAGGCT TCTGGAGATG

          SalI
          ~~~~~
 AscI
          ~~~~~
1051 AGCCGTTTAT TCGAAATGGC CACTATCTCC CAAGCTGCCCT ACGCCGACCT GTGCAACATTT CCGTCGACTA
    TCGGCAAAATC AGCTTTTACC GTGATAGAGG GTTCGACGGA TCGCGGCTGA CACGTTGTAA GGCAGCTGAT

          BamHI
          ~~~~~
1121 TTATCAAGGG AGAGAAAATT TACAATTCTC AAAGTGACAT TAACGATGG ATCCTCCGGC ACGACAGCAG
 AATAGTTCCC TCCTCTTTAA ATGTTAAGAG TTTGACTGTA ATTGCTTACC TAGGAGGCGC TGCTGTCTGC

```

FIG. 42C

```

1191 CAAAGAAATA ATCACCGTCT TCCGTGGCAC TGGTAGTGAT ACGAATCTAC AACTCGATAC TAACTACACC
 GTTCTTTTAT TAGTGGCAGA AGGCACCGTG ACCATCACTA TGCTTAGATG TTGAGCTATG ATTGATGTGG

1261 CTCACGCCCTT TCGACACCCCT ACCACAATGC AACGGTTTGG AAGTACACGG TGGATATTAT ATTGGATGGG
 GAGTGGGGAA AGCTGTGGGA TGGTGTACG TTGCCAACAC TTTCATGTGCC ACCATAAATA TAACTTACCC

1331 TCTCCGTCCA GGACCAAGTC GAGTCGCTTG TCAAAACAGCA GGTAGCCAG TATCCGGACT ACGCGCTGAC
 AGAGGCAGGT CCTGGTTTCAG CTCAGCGAAC AGTTTGTCTGT CCAATCGGTC ATAGGCTGA TCGCGGACTG

1401 CGTGACCGGC CACKCCCTCG GCGCTTCCCT GCGGGCACTC ACTGCCGCC AGCTGTCTGC GACATACGAC
 GCACTGGCCG GTGMMGGGAGC CCGGGAGGA CCGCCGTGAG TGACGGCGGG TCGACAGACG CTGTATGCTG

1471 AACATCCGCC TGTACACCTT CCGGGAACC CGCAGCGGCA ATCAGGCCCT CCGCTCGTAC ATGAACGATG
 TTGTAGGCGG ACATGTGGAA GCCGCTTGGC CGCTCGCCGT TAGTCCGGAA GCGCAGCATG TACTTGCTAC

 XhoI
                                ~~~~~~

1541 CCTTCCAAGC CTCGAGCCCA GATACGACGC AGTATTTCCG GGTCACTCAT GCCAACGACG GCATCCCAA
    GGAAGGTTCT GAGCTCGGGT CTATGCTGCG TCATAAAGGC CCAGTGAGTA CGGTGCTGC CGTAGGGTTT

                                NcoI
                                ~~~~~~

1611 CCTGCCCCCG GTGGAGCAGG GGTACGCCCA TGGCGGTGTA GAGTACTGGA GCGTTGATCC TTACAGCGCC
 GGACGGGGGC CACCTCGTCC CCATGCGGGT ACCGCCACAT CTCATGACCT CGCAACTAGG AATGTCGCGG

1681 CAGAACACAT TTGTCTGCAC TGGGGATGAA GTGCAGTCT GTGAGGCCCA GGGCGGACAG GGTGTGAATA
 GTCTTGTGTA AACAGACGTG ACCCTACTT CACGTCACGA CACTCCGGGT CCGCCTGTC CCACACTTAT

1751 ATGCGCACAC GACTTATTTT GGGATGACGA GCGGAGCCCTG TACATGCTGA TCAGTCATTT CAGCCTCCCC
 TACGCGTGTG CTGAATATAA CCCTACTGCT CCGCTCGGAC ATGTACCACT AGTCAGTAAA GTCGGAGGGG

1821 GAGTGTACCA GGAAGATGG ATGTCCTGGA GAGGGGGCCG CGTAACCACT GAAGGATGAG CTGTAAAGAA
 CTCACATGGT CCTTCTTACC TACAGGACCT CTCCCCCGGC GCATTTGGTA CTTCCTACTC GACATTTCTT

```

**FIG. 42D**

```

1891 GCAGATCGTT CAAACATTGG GCAATAAAGT TTCTTAAGAT TGAATCCTGT TGC CGGTCTT GC GATGATTA
CGTCTAGCAA GTTTGTAAAC CGTTATTCA AAGAATTCTA ACTTAGGACA ACGCCAGAA CGCTACTAAT

1961 TCATATAAAT TCTGTGGAAT TACGTTAAGC ATGTAATAAT TAACATGTAA TGCATGACGT TATTATGAG
AGTATATATA AGACAACCTA ATGCAATTGG TACATTATTA ATTGTACATT ACGTACTGCA ATAAATACTC

BssHII
      ~~~~~
2031 ATGGGTTTTT ATGATTAGAG TCCCGCAATT ATACATTTAA TACGGGATAG AAAACAAAAT ATAGCGCGCA
TACCCAAAAA TACTAATCTC AGGCGGTAA TATGTAAATT ATGGGCTATC TTTTGTTTTA TATCGCGCGT

BssHII
      ~~~~~
2101 AACTAGGATA AATTATCGCG CGCGGTGTCA TCTATGTTAC TAGATCGATA AGCTTCTAGA GCGGCCGGTG
TTGATCCCTAT TTAATAGCGC GCGCCACAGT AGATACAATG ATCTAGCTAT TCGAAGATCT GCGCGGCCAC

BssHII
      ~~~~~
2171 GAGCTCCAAT TCGCCCTATA GTGAGTCGTA TTACGCGCGC TCACGTGGCG TCGTTTACA ACGTCGTGAC
CTCGAGGTTA AGCGGGATAT CACTCAGCAT AATGCGCGCG AGTGACCGGC AGCAAAATGT TGCAGCACTG

2241 TGGGAAAACC CTGGCGTTAC CCAACTTAAT CGCCTTGCAG CACATCCCCC TTTCGCCAGC TGGCGTAATA
ACCTTTTGG GACCGCAATG GGTGAATTA GCGGAACGTC GTGTAGGGGG AAGCGGTG ACCGCATTAT

2311 GCGAAGAGGC CCGCACCGAT CGCCTTCCC AACAGTTGCG CAGCCTGAT GCGAATGGG ACGCGCCCTG
CGCTTCTCCG GCGGTGGCTA GCGGGAAGGG TTGTCAACGC GTCGGACTTA CCGCTTACCC TCGCGCGGAC

2381 TAGCGGCGCA TTAAGCGCGG CGGTGTGGT GGTACGCGC AGCGTGACC CTACACTTGC CAGCGCCCTA
ATCGCCGCGT AATTCGCGCC GCCACACCA CCAATGCGCG TCGCACTGGC GATGTGAACG GTTCGCGGGAT

2451 GCGCCCGCTC CTTTCGCTTT CTTCCCTTCC TTCTTCGCCA CGTTCGCGCG CTTTCCCGGT CAAGCTCTAA
CGCGGGCGAG GAAAGCGAAA GAAGGAAGG AAGAGCGGT GCAAGCGGCC GAAAGGGGCA GTTCGAGATT

```

XbaI  
~~~~~

Clai HindIII
~~~~~

BssHII  
~~~~~

BssHII
~~~~~

FIG. 42E

2521 ATCGGGGGCT CCCTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC CCCAAAAAAC TTGATTAGGG  
 TAGCCCCCGA GGGAAATCCC AAGGTAAAT CACGAANTGC CGTGGAGCTG GGGTTTTTTG AACTAATCCC  
 2591 TGATGGTTCA CGTAGTGGC CATCGCCCTG ATAGACGGTT TTTCGCCCTT TGACCTTGA GTCCACGTTT  
 ACTACCAAGT GCATCACCCG GTAGCGGAC TATCTGCCAA AAGCGGGA ACTGCAACCT CAGGTGCAAG  
 2661 TTTAATAGTG GACTCTTGT CCAAACTGGA ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTAT  
 AAATTATCAC CTGAGAACAA GGTGTGACCT TGTGTGAGT TGGGATAGAG CCAGATAAGA AACTAAATA  
 2731 AAGGATTTT GCCGATTTCG GCCTATTGGT TAAAAAATGA GCTGATTAA CAAAAATTA ACGCGAATT  
 TTCCCTAAAA CGGCTAAAGC CGGATAACCA ATTTTACT CGACTAAAT GTTTTAAAT TGGCTTAA  
 2801 TAACAAAATA TTAACGCTTA CAATTTAGGT GGCACTTTC GGGGAAATGT GCGCGCTGG CCTATTGTT  
 ATTGTTTAT ARTTGCGAAT GTTAAATCCA CCGTGAAAG CCCCTTTACA CCGCGCTGG GGTAAACAA  
 2871 TATTTTCTA AATACATTCA AATATGTATC CGTCTATGAG ACAATAACCC TGAATAATGC TTCAATAATA  
 ATAAAAAGAT TTATGTAGT TTATACATAG GCGAGTACT TGTATTGGG ACTATTACG AAGTTATTAT  
 2941 TTGAAAAAGG AAGAGTATGA GTATTCAACA TTTCCTGTC GCCCTTATTC CCTTTTTCG GGCATTTTC  
 AACTTTTTCC TTCTCATACT CATAAGTTGT AAAGGCACAG CGGGAATAAG GGAATAAACG CCGTAAACG  
 3011 CTTCCCTGTTT TTGCTCACCC AGAAGCGCTG GTGAAAGTAA AAGATGCTGA AGATCAGTTG GGTGCACGAG  
 GAAGGACAAA AACGAGTGGG TCTTTGCGAC CACTTTCATT TTCTACGACT TCTAGTCAAC CCACGTGCTC  
 3081 TGGGTTACAT CGAACTGGAT CTCAACAGCG GTAAAGATCCT TGAGAGTTTT CGCCCCGAAG AACGTTTCC  
 ACCCAATGTA GCTTGACCCTA GAGTTGTGCG CATCTAGGA ACTCTCAAAA CCGGGCTTC TTGCAAAAGG  
 3151 AATGATGAGC ACTTTTAAAG TTCTGCTATG TGGCGCGGTA TTATCCCGTA TTGACGCCGG GCAAGAGCAA  
 TTACTACTCG TGAAAAATTC AAGACGATAC ACCGCGCCAT AATAGGGCAT AACTGCGGCC CGTTCTCGTT  
 3221 CTCGGTCGCC GCATACACTA TTCTCAGAAAT GACTTGGTTG AGTACTCAC AGTCACAGAA AAGCATCTTA  
 GAGCCAGCGG CGTATGTGAT AAGAGTCTTA CTGAACCAAC TCATGAGTGG TCAGTGTCTT TTCGTAGNAT  
 3291 CGGATGGCAT GACAGTAAGA GAATTATGCA GTGCTGCCAT AACCATGAGT GATAACACTG CGGCCAACTT  
 GCCTACCGTA CTGTCAATTCT CTTAATACGT CACGACGGTA TTGGTACTCA CTATTGTGAC GCCGGTTGAA

**FIG. 42F**

3361 ACTTCTGACA ACGATCGGAG GACCGAAGGA GCTAACCGCT TTTTTCACAA ACATGGGGGA TCATGTAAC T  
 TGAAGACTGT TGCTAGCCTC CTGGCTTCCT CGATTGGCGA AAAAACGTGT TGTACCCCTT AGTACATTGA  
 3431 CGCCTTGATC GTTGGGAACC GGAGCTGAAT GAAGCCATAC CAAACGACGA GCGTGACACC ACGATGCCCTG  
 GCGGAAC TAG CAACCCCTGG CCTCGACTTA CTTTCGGTATG GTTTGCTGCT CGCACTGTGG TGCTACGGAC  
 3501 TAGCAATGGC AACAACTGT CGCAAACTAT TAACTGGCGA ACTACTTACT CTAGCTTCCC GGCAACAAT  
 ATCGTTACCG TTGTTGCAAC GCGTTGATA ATTGACCGCT TGATGAATGA GATCGAAGGG CCGTTGTAA  
 3571 AATAGACTGG ATGGAGGCGG ATAAAGTTC ATAAAGTTC AGGACCACTT CTGCGCTCGG CCTTCCGGC TGGCTGGTT  
 TTATCTGACC TACCTCCGCC TATTTCAACG TCCTGGTGA GACGCGAGCC GGGAGGGCG ACCGACCAA  
 3641 ATTGCTGATA AATCTGGAGC CGGTGAGCGT GGTCTCGCG GTATCATTC AGCACTGGG CCAGATGGTA  
 TAACGACTAT TTAGACCTCG GCCACTCGCA CCCAGAGCG CATAGTAAAC TCGTGACCCC GGTCTACCAT  
 3711 AGCCCTCCCG TATCGTAGTT ATCTACACGA CGGGGAGTCA GGCAACTATG GATGAACGAA ATAGACAGAT  
 TCGGGAGGGC ATAGCATCAA TAGATGTGCT GCCCCTCAGT CCGTTGATAC CTACTTGTCTT TATCTGTCTA  
 3781 CGCTGAGATA GGTGCCCTAC TGATTAAGCA TTGGTAACTG TCAGACCAAG TTTACTCATA TATACTTTAG  
 GCGACTCTAT CCACGGAGTG ACTAATTCGT AACCATTTGAC AGTCTGGTTC AAATGAGTAT ATATGAAATC  
 3851 ATTGATTAA AACTTCAATTT TTAATTAAA AGGATCTAGG TGAAGATCCT TTTTGTATAAT CTCATGACCA  
 TAACTAAATTT TTGAAGTAAA AATTAAATTT TCCTAGATCC ACTTCTAGGA AAAACTATTA GAGTACTGGT  
 3921 AATCCCTTA ACGTGAGTTT TCGTCCACT GAGGCTCAGA CCGGCTAGAA AAGATCAAAG GATCTTCTTG  
 TTTAGGGAAT TGCACCTCAA AGCAAGGTGA CTCGCAGTCT GGGGCTCTT TTCTAGTTTC CTAGAAGAAC  
 3991 AGATCCCTTTT TTTCTGCGCG TAATCTGCTG CTTGCAAAACA AAAAACCCAC CGCTACCAGC GGTGGTTTGT  
 TCTAGGAAAA AAAGACGCGC ATTAGACGAC GAACGTTTGT TTTTGTGGTG GCGATGGTTC CCACCAACA  
 4061 TTGCCGGATC AAGAGCTACC AACTCTTTT CCGAAGGTAA CTGSCCTCAG CAGAGCGCAG ATACCAATA  
 AACGGCCTAG TTCTCGATGG TTGAGAAAAA GGTCTCCATT GACCGAAGTC GTCTCGCGTC TATGGTTTAT  
 4131 CTGTCTTCT AGTGTAGCCG TAGTTAGGCC ACCACTCAA GAACCTGTA GCACCGCCTA CATACCTCGC  
 GACAGGAAGA TCACATCGGC ATCAATCCGG TGGTGAAGTT CTTGAGACAT CGTGGCGGAT GTATGGAGCG

**FIG. 42G**



4201 TCTGCTAATC CTGTTACCAG TGGCTGCTGC CAGTGGCGAT AAGTCGTGTC TTACCGGGTT GGACTCAAGA  
AGACGATTAG GACAATGGTC ACCGACGACG GTCACCGCTA TTCAGCACAG AATGGCCCAA CCTGAGTTCT

4271 CGATAGTTAC CGGATAAGGC GCAGCGGTGC GGTGAACGG GGGTTTCGTG CACACAGCCC AGCTTGGAGC  
GCTATCAATG GCCTATTCCG CGTCGCCAGC CCGACTTGCC CCCCAAGCAC GTGTGTGCGG TCGAACCTCG

4341 GAACGACCTA CACCGAATG AGATACCTAC AGCTGAGCT ATGAGAAAAG GCCACGCTTC CCGAAGGGAG  
CTTGCTGGAT GTGGCTTGAC TCTATGGATG TCGCACTCGA TACTCTTTTCG CGGTGCGAAG GGCTTCCCTC

4411 AAAGGCGGAC AGGTATCCGG TAAGCGGCAG GGTGGAACA GGAGAGCGCA CGAGGGAGCT TCCAGGGGGA  
TTTCCGCCCTG TCCATAGGCC ATTCCGCCGC CCAGCTTGT CCTCTCGCGT GCTCCCTCGA AGTCCCCCT

4481 AACGCTGGT ATCTTTATAG TCCTGTGCGG TTTCCGCCACC TCTGACTTGA GCCTCGATT TTTGTGATGCT  
TTGCGGACCA TAGAATATC AGGACAGCCC AAAGCGGTGG AGACTGAACT CGCAGCTAAA AACACTACGA

4551 CGTCAGGGG GCGGAGCCTA TGGAAAAACG CCAGCAACGC GGCTTTTITA CGGTTCCTGG CCTTTTGTG  
GCAGTCCCC CGCTTCGGAT ACCTTTTTCG GGTGTTGCG CCGGAAAAAT GCCAAGGACC GGAAAAACGAC

4621 GCCTTTTGT CACATGTTCT TTCCTGGGTT ATCCCTGAT TCTGTGATA ACCGTATTAC CGCTTTTGG  
CGGAAAAACGA GTGTACAAGA AAGGACGCAA TAGGGACTA AGACACCTAT TGGCATATG GCGGAAACTC

4691 TGAGCTGATA CCGCTCGCCG CAGCCGAACG ACCGAGCGCA GCGAGTCACT GAGCGAGGAA GCGGAAGAGC  
ACTCGACTAT GCGGAGCGGC GTCGGCTTGC TGGCTCGCGT CGCTCAGTCA CTCGCTCCTT CGCCTTCTCG

4761 GCCCAATACG CAAACCGCCT CTCGCCGCGC GTTGGCCGAT TCATTATGC AGCTGGCAGC ACAGGTTTCC  
CGGTTATATG GTTTGGCGGA GAGGGGCGCG CAACCGGCTA AGTAATFACG TCGACCGTGC TGTCCAAAGG

4831 CGACTGGAAA GCGGGCAGTG AGCGCAACGC AATTAATGTG AGTTAGCTCA CTCATTAGGC ACCCCAGGCT  
GCTGACCTTT CGCCCGTCAC TCGCGTTGCG TTAATTACAC TCAATCGAGT GAGTAATCCG TGGGGTCCGA

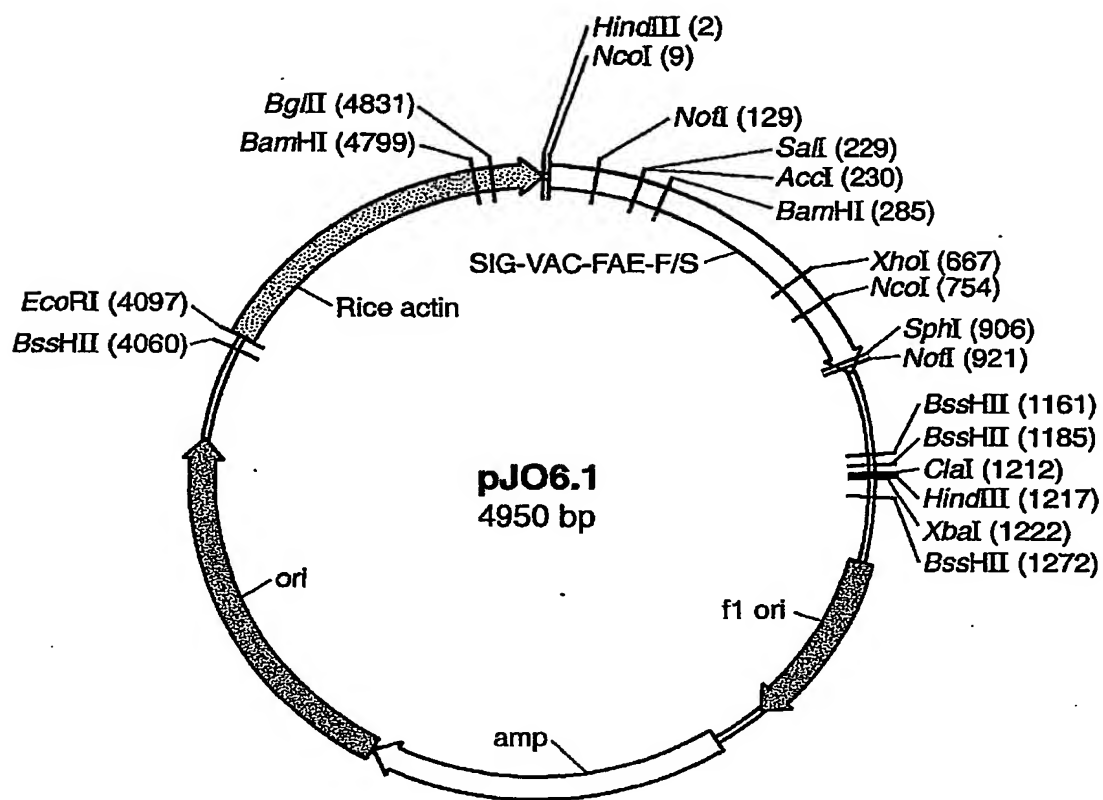
4901 TTACACTTTA TGCTTCCGGC TCGTATGTTG TGTGGAATTG TGAGCGGATA ACAATTTCAC ACAGGAAACA  
AATGTGAAT ACGAAGGCCG AGCATAACA ACACCTTAAC ACTCGCCTAT TGTAAAGTG TGTCTTTTGT

FIG.\_42H

4971

BSSHII	NGOI
~~~~~	Kpni
~~~~~	~~~~~
GCTATGACCA TGAATTACGCC AAGCGCGCAA TTAACCCCTCA CTAAGGGA CAAAGCTGG GTAC	
CGAFACTGGT ACTAATGCCG TTCGCGCGTT AATGGGAGT GATTTCCCTT GTTTCGACC CATG	

FIG.\_42I

**FIG. 43A**

HindIII NcoI  
 ~~~~~  
 1 AAGCTTACCA TGGCCACGC CCGCTCCTC CTCTGGCG TCGCCGTGCT GGCCACGGC GCCGTGCGC
 TTGGAATGGT ACCGGGTGG GGCACAGAG GAGGACCGG AGCGGCACGA CCGGTGCCG CGGCAGCGC

NotI
 ~~~~~  
 71 TCGCCTCCTC CTCTCCTTC GCCACTCCA ACCGATCCG GCCGTGACC GGCCGCGCG CGCCTCCAC  
 AGCGAGGAG GAGGAGGAG CGCTGAGGT TGGGTAGGC CGGCAGTGG CCGCGCGCC GCGGAGGTG

141 GCAGGGCATC TCCGAAGACC TCTACAGCC TTTAGTCGA ATGGCCACTA TCTCCCAAG TGCCTACGCC  
 CGTCCCGTAG AGGCTCTGG AGATGTCGG AAATCAGCTT TACCGGTGAT AGAGGTTG ACGGATGCG

Sall  
 ~~~~~  
 211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTACG
 CTGGACACGT TGTAAAGCAG CTGATAATAG TTCCCTCTCT TTTAAATGTT AAGAGTTTGA CTGTAAATGC

BamHI
 ~~~~~  
 281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AAATAATCAC CGTCTTCCGT GGCACTGGTA GTGATACGA  
 CTACCTAGGA GCGCTGCTG TCGTCGTTTC TTTATTATAG GCAGAGGCA CCGTGACCAT CACTATGCTT

351 TCTACAACTC GATACTAACT ACACCTCAC GCCTTTCGAC ACCCTACCAC AATGCAACGG TTGTGAGTA  
 AGATGTTGAG CTATGATGA TGTGGGAGTG CGGAAAGCTG TGGGATGGTG TTACGTTGCC AACACTTCAT

421 CACGGTGGAT ATTATATTGG ATGGGTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA  
 GTGCCACCTA TAATATAACC TACCCAGAGG CAGGTCCTGG TTCAGCTCAG CGAACAGTTT GTCGTCCAAT

491 GCCAGTATCC GGACTACGG CTGACCGTGA CCGGCCACKC CCTCGGCGCC TCCCTGGCG CACTACTGC  
 CCGTCATAGG CCTGATGCG GACTGGCACT GCGCGGTGAG GAGCCGCGG AGGACCGCC GTGAGTGACG

561 CGCCAGCTG TCTGCGACAT ACGACAACAT CGGCTGTAC ACCTTCGGCG ACCCGGCG CCGCAATCAG  
 GCGGTCGAC AGACGCTGTA TGCTGTTGTA GCGGACATG TGAAGCCGC TTGGCGCGTC GCCGTTAGTC

**FIG. 43B**

XhoI  
~~~~~

631 GCCTTCGCGT CGTACATGAA CGATGCCCTTC CAAAGCTCGA GCCAGATAC GACGCAGTAT TTCCGGGTCA
CGGAAGCGCA GCATGTACTT GCTACGGAAG GTTCGGAGCT CGGCTCTATG CTGCGTCATA AAGGCCAGT

NcoI
~~~~~

701 CTCATGCCAA CGACGGCATC CCAAACCTGC CCGCGGTGGA GCAGGGGTAC GCCCATGGCG GTGTAGAGTA  
GAGTACGGTT GCTGCCGTAG GGTTTGGACG GGGGCCACCT CGTCCCCATG CCGGTACCGC CACATCTCAT

771 CTGGAGCGTT GATCCTTACA GCGCCCAGAA CACATTTGTC TGCACCTGGG ATGAAGTGCA GTGCTGTGAG  
GACCTCGCAA CTAGGAATGT CGCGGGTCTT GTGTAAACAG ACGTGACCCC TACTTCACGT CACGACACTC

SphI  
~~~~~

841 GCCCAGGGCG GACAGGGTGT GAATAATGCG CACACGACTT ATTTGGGAT GACGAGCGGC GCATGCACCT
CGGTCCCGC CTGTCCACCA CTTATTACGC GTGTCTGAA TAAACCCCTA CTGCTCGCG CGTACGTGGA

NotI
~~~~~

911 GGCCGGTCCG GGCCGCGGAA ACCACTGAAG GATGAGCTGT AAAGAAGCAG ATCGTTCAA CATTGGCAA  
CCGCCAGCG CCGCGCCCTT TGGTGACTTC CTACTCGACA TTCTTTCGTC TAGCAAGTTT GTAAACCGTT

981 TAAAGTTTCT TAAGATTGAA TCCTGTTGCC GGTCTTGCAG TGATTATCAT ATAAATTTCTG TTGAATTACG  
ATTTCAAAGA ATCTAACTT AGGACAACGG CCAGAACGCT ACTAATAGTA TATTAAGAC AACTTAATGC

1051 TTAAGCATGT AATAATTAC ATGTAATGCA TGACGTTATT TATGAGATGG GTTTTATGA TTAGAGTCCC  
AATTGGTACA TTATTAAATG TACATTACGT ACTGCAATAA ATACTCTACC CAAAAFACT AATCTCAGGG

BssHII  
~~~~~

1121 GCAATTATAC ATTAAATACG CGATAGAAA CAAAATATAG CGCGCAACT AGGATAAAT ATCGCGCGCG
CGTTAAATATG TAAATTATGC GCTATCTTTT GTTTTATATC GCGCGTTTGA TCCTATTAA TAGCGCGCGC

BssHII
~~~~~

FIG.\_43C

XbaI

Clai HindIII

1191 GTGTCTATCTA TGTACTAGA TCGATAAGCT TCTAGAGCGG CCGGTGGAGC TCCAAATTCG CCTATAGTGA  
CACAGTAGAT ACAATGATCT AGCTATTGCA AGATCTCGCC GGCACACTCG AGGTTAAGCG GGATATCACT

BssHII

1261 GTCTGATTAC GCGCGCTCAC TGGCCGTCGT TTTACAACGT CGTGACTGGG AAAACCCCTGG CGTTACCCAA  
CAGCATAAATG CCGCGGAGTG ACCGGCAGCA AAATGTTGCA GCACTGACCC TTTTGGGACC GCAATGGGTT

1331 CTTAATCGCC TTGCAGCACA TCCCCCTTTC GCCAGCTGGC GTAATAGCGA AGAGGCCCGC ACCGATCGCC  
GAATTAGCGG AACGTCGTGT AGGGGGAAG CCGTCGACCG CATATATCGT TCTCCGSGCG TGGCTAGCGG

1401 CTTCCCAACA GTTGGCGAGC CTGAATGGCG AATGGGACGC GCCCTGTAGC GCGGCATTAA GCGCGGCGGG  
GAAGGGTTGT CAACGCGTCG GACTTACCGC TTACCTCTCG CCGGACATCG CCGCGTAATT CCGCGCGCCC

1471 TGTGGTGGTT ACGCGCAGCG TGACCGCTAC ACTTGCAGC GCCCTAGCGC CCGCTCCCTTT CGCTTCTTTC  
ACACCACCAA TCGCGGTCGC ACTGGCGATG TGAACGGTCG CCGGATCGCG GCGGAGGAAA GCGAAAGAAAG

1541 CCTTCCCTTC TCGCCACGTT CCGCGGCTTT CCGCGTCAAG CTCTAAATCG GGGGCTCCCT TTAGGGTTCC  
GGAAGGAAAG AGCGGTGCAA GCGGCCGAAA GGGGCACTTC GAGATTAGC CCGCGAGGGA AATCCCAAGG

1611 GATTTAGTGC TTTACGGCAC CTCGACCCCA AAAAACTTGA TTAGGGTGAT GGTTCACGTA GTGGGCCATC  
CTAAATCAGG AATGCCGTG GAGCTGGGGT TTTTGAAC TATCCACTA CCAAGTGCAT CACCCGGTAG

1681 GCCCTGATAG ACGGTTTTC GCCCTTTGAC GTTGGAGTCC ACGTCTTTA ATAGTGGACT CTTGTTCCAA  
CGGGACTATC TGCCAAAAAG CCGGAAACTG CAACCTCAGG TGCAAGAAAT TATCACCTGA GAACAAGGTT

1751 ACTGGAACAA CACTCAACCC TATCTCGGTC TATTCTTTTG ATTTATAAGG GATTTTGCCG ATTTGGGCTT  
TGACCTTGTT GTGAGTTGGG ATAGAGCCAG ATAGAAAAAC TAAATATTCC CTAATAACGC TAAAGCCGGA

1821 ATTGGTTAAA AAATGAGCTG ATTTAACAAA AATTTAACGC GAATTTTAAC AAAATATTAA CGCTTACAAT  
TAACCAATTT TTTACTCGAC TAAATTTGTT TTAATTTGCG CTTAAAAATTG TTTTATAATT GCGAATGTTA

FIG. 43D

1891 TTAGGTGGCA CTTTTCGGG AAATGTGCGG GGAACCCCTA TTTGTTTATT TTTCTAAATA CATTCAATA  
 AATCCACCGT GAAAGCCCC TTACACGCG CCTTGGGAT AACAAATAA AAAGATTAT GTAAGTTAT  
 1961 TGTATCCGCT CATGAGACAA TAACCCCTGAT AAATGCTTCA ATAATATTGA AAAAGGAAGA GTATGAGTAT  
 ACATAGGCGA GTACTCTGTT ATTGGACTA TTACGAAAT TATTATAACT TTTTCCCTCT CATACTCATA  
 2031 TCAACATTTC CGTGTGCCCC TTATTCCCTT TTATGCGGCA TTTTGCCTTC CTGTTTTTGC TCACCCAGAA  
 AGTTGTAAAG GCACAGCGG AATAAGGGA AAAACGCCGT AAAACGGAAG GACAAAAACG AGTGGGTCTT  
 2101 ACGTGGTGA AAGTAAAGA TGCTGAAGAT CAGTTGGTG CACGAGTGG TTACATCGAA CTGGATCTCA  
 TCGACCACT TTCATTTTCT ACGACTTCTA GTCAACCCAC GTGCTCACCC AATGTAGCTT GACCTAGAGT  
 2171 ACAGCGGTAA GATCCTTGAG AGTTTTCGCC CCGAAGAAG GGTCTCTTGC AAAAGGTTAC TACTCGTGAA AATTTCAAGA  
 TGTGCGCCATT CTAGGAACTC TCAAAAGCGG GGTCTCTTGC AAAAGGTTAC TACTCGTGAA AATTTCAAGA  
 2241 GCTATGTGGC GCGGTATTAT CCGGTATTGA CCGCGGGCA GAGCAACTCG GTCGCCGCGT ACCTATTCTT  
 CGATACACCG CGCCATAATA GGGCATAACT GCGGCCCGTT CTCGTTGAGC CAGCGGCGTA TGTGATAAGA  
 2311 CAGATGACT TGGTTGAGTA CTCACCAGTC ACAGAAAGC ATCTTACGGA TGGCATGACA GTAGAGAAAT  
 GTCTTACTGA ACCAACTCAT GAGTGGTCAG TGTCTTTTCG TAGAATGCCT ACCGTACTGT CATTCTCTTA  
 2381 TATGCAGTGC TGCCATAACC ATGAGTGATA ACCTGCGGC CAACTTACTT CTGACAACGA TCGGAGGACC  
 ATAGGTCACG ACGGTATTGG TACTCACTAT TGTGACGCG GTTGAATGAA GACTGTTGCT AGCTCCTGG  
 2451 GAAGGAGCTA ACCGCTTTT TGCACAACAT GGGGATCAT GTAACTCGCC TTGATCGTTG GGAACCGGAG  
 CTTCTCTGAT TGGCGAAAA ACGTGTGTA CCCCCTAGTA CATTGAGCG AACTAGCAAC CCTTGGGCTC  
 2521 CTGAATGAAG CCATACCAAA CGACGAGCGT GACACACGA TGCCTGTAGC AATGGCAACA ACGTTGCGCA  
 GACTTACTTC GGTATGGTTT GCTGCTCGCA CTGTGTGCTT ACGGACATCG TTACCGTTGT TGCACGCGT  
 2591 AACTATTAACT TGGCGAACTA CTTACTCTAG CTTCCCGGCA ACAATTAATA GACTGGATGG AGGCGGATAA  
 TTGATAATTG ACCGCTTGAT GAATGAGATC GAAGGCGCGT TGTAAATTAT CTGACCTACC TCCGCCATAT  
 2661 AGTTGCAGGA CCACCTCTGC GCTCGGCCCT TCCGGCTGG TGGTTTATTG CTGATAAATC TGGAGCCCGT  
 TCAACGTCCT GGTGAAGACG CGAGCCGGA AGGCCGCGG ACCAAATAAC GACTATTAG ACCTCGCGCA

**FIG. 43E**

2731 GAGCGTGGGT CTCGCGGTAT CATTGCAGCA CTGGGGCCAG ATGGTAAGCC CTCCGGTATC GTAGTTATCT  
CTCGCACCCA GAGCGCCATA GTAACTGCTGT GACCCCGGTC TACCATTCCG GAGGCATAG CATCAATAGA

2801 ACACGACGGG GAGTCAGGCA ACTATGGATG AACGAATAG ACAGATCGCT GAGATAGGTG CCTCACTGAT  
TGTGCTGCCC CTCAGTCCGT TGATACCTAC TTGCTTTATC TGTCTAGCGA CTCTATCCAC GGAGTGACTA

2871 TAAGCATTTG TAACTGTGAG ACCAAGTTTA CTCATATATA CTTTAGATTG ATTTAAAACT TCATTTTTTAA  
ATTGCTAACC ATTGACAGTC TGGTTCAAT GAGTATATAT GAAATCTAAC TAAATTTTGA AGTAAAAATT

2941 TTTAAAAAGG TCTAGGTGAA GATCCTTTTT GATAATCTCA TGACCAAAAT CCTTTAAAGT GAGTTTTCGT  
AAATTTTCTT AGATCCACTT CTAGGAAAAA CTATTAGAGT ACTGGTTTTA GGAATTTGCA CTCAAAAAGCA

3011 TCCACTGAGC GTCAGACCCC GTAGAAAAA TCAAAGGATC TTCTTGAGAT CCTTTTTCCT TCGCGGTAAT  
AGGTGACTCG CAGTCTGGG CATCTTTTCT AGTTTCTTAG AAGAACTCTA GGAATAAAG ACGGGCATTA

3081 CTGCTGCTTG CAAACAAAAA AACCAACGCT ACCAGCGGTG GTTTGTTTGC CGGATCAAGA GCTACCAACT  
GACGACGAAC GTTTGTTTTT TTGGTGGCGA TGGTCGCCAC CAAACAAACG GCCTAGTTCT CGATGGTTGA

3151 CTTTTCCTGA AGGTAACCTG CTTCAGCAGA GCGCAGATAC CAAATACGT CCTTCTAGTG TAGCCGTAGT  
GAAAAAGGCT TCCATTGACC GAAGTCGTCT CCGCTCTATG GTTTATGACA GGAAGATCAC ATCGGCATCA

3221 TAGGCCACCA CTTCAAGAAC TCTGTAGCAC CGCTACATA CCTCGCTCTG CTAATCCTGT TACCAGTGGC  
ATCCGGTGGT GAAGTTCTTG AGACATCGTG GCGGATGTAT GGAGCGAGAC GATTAGGACA ATGGTCACCG

3291 TGCTGCCAGT GCGGATAAGT CGTGTCTTAC CCGGTTGGAC TCAAGACGAT AGTTACCGGA TAAGGCGCAG  
ACGACGGTCA CCGCTATTCA GCACAGAATG GCCCAACCTG AGTTCTGCTA TCAATGGCCCT ATTCCGCGTC

3361 CCGTCGGGCT GAACGGGGG TTGCTGCACA CAGCCAGCT TGGAGCGAAC GACCTACACC GAACTGAGAT  
GCCAGCCCGA CTTGCCCCCC AAGCACGTGT GTGCGGTGGA ACCTCGCTTG CTGGATGTGG CTTGACTCTA

3431 ACCTACAGCG TGAGCTATGA GAAAGCGCCA CGCTTCCCGA AGGAGAAAG GCGGACAGGT ATCCGGTAAG  
TGGATGTCGC ACTCGATACT CTTTCGCGGT GCGAAGGCT TCCCTCTTTC CGCCTGTCCA TAGGCCATTC

3501 GGGCAGGGTC GGAACAGGAG AGCGCACGAG GGAGTTCCA GGGGAAACG CCTGGTATCT TTATAGTCTT  
GCCGTCCAG CCTTGTCCTC TCGCGTGTCT CCTGAAGGT CCCCCTTTCG GGACCATAGA AATATCAGGA

**FIG.\_43F**

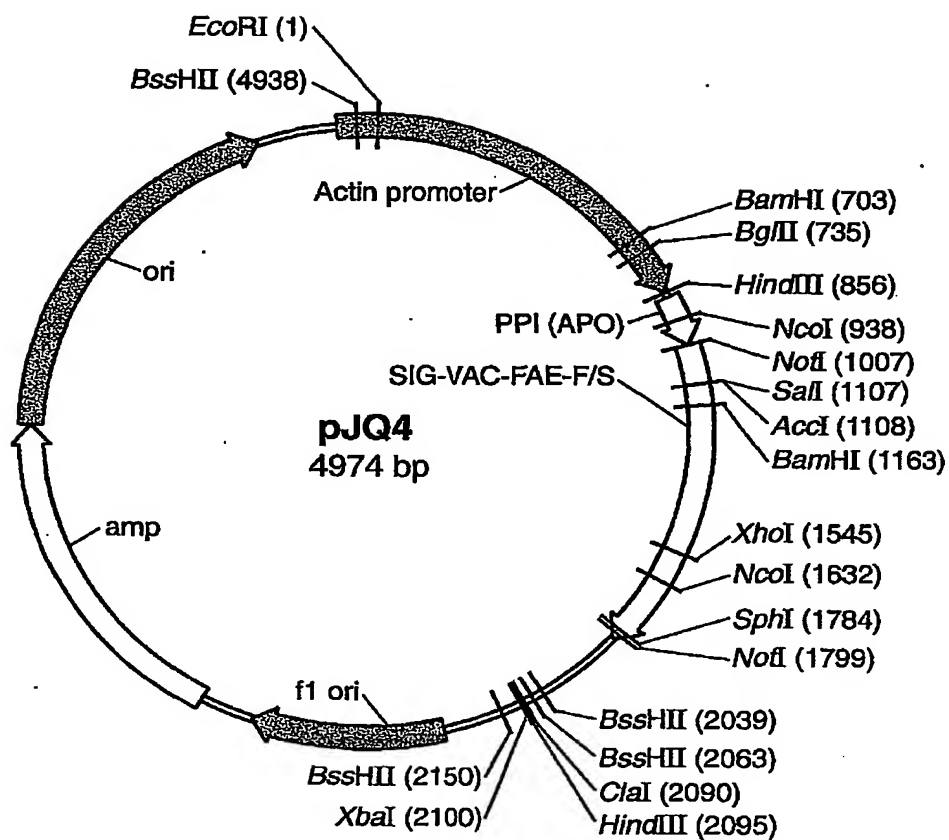


3571 GTCGGGTTTC GCCACCTCTG ACTTGAGCGT CGATTTTGT GATGTCGTC AGGGGGCGG AGCCTATGGA  
 CAGCCCAAAG CGGTGGAGAC TGAACCTGCA GCTAAACA CTAAGAGCAG TCCCCCGCC TCGATACCT  
 3641 AAAACGCCAG CAACGCGGCC TTTTACGGT TCCTGGCCCT TTTGCTCACA TGTCTTTCC  
 TTTTGGCGTC GTTGGCGCGG AAAAATGCCA AGGACCGGA AACGACGGA AACGAGTGT ACAAGAAAGG  
 3711 TGGCTTATCC CCTGATTCTG TGGATAACCG TATTACCGCC TTTGAGTGAG CTGATACCGC TCGCCGCGAGC  
 ACGCAATAGG GGAATAAGAC ACCTATTGGC ATAATGGCGG AAACCTACTC GACTATGGCG AGCGGCGTCCG  
 3781 CGAACGACCG AGCGCAGCGA GTCAGTGAGC GAGGAAGCGG AAGAGCGCC AATACGCAA ACGCCTCTCC  
 GCTTGCTGGC TCGCGTCCGT CAGTCACTCG CTCCTTCGCC TTCTCGCGG TTATGCGTTT GCGGAGAGG  
 3851 CCGCGCGTTG GCCGATTCTAT TAATGCAGCT GGCACGACAG GTTCCCGAC TGGAAAGCGG GCAGTGAGCG  
 GCGCGCGCAAC CCGCTAAGTA ATTACGTCTGA CCGTGTCTGC CAAAGGCTG ACCTTTCGCC CGTCACTCGC  
 3921 CAACGCAATT AATGTAGTT AGCTCACTCA TTAGGCACCC CAGGCTTAC ACTTTATGCT TCCGGCTCGT  
 GTTGCCTTAA TTACACTCAA TCGAGTGAGT AATCCGTGGG GTCCGAAATG TGAATACGA AGGCCGAGCA  
 BssHII  
 3991 ATGTTGTGTG GAATTGTGAG CGGATAACAA TTTCACACAG GAAACAGCTA TGACCATGAT TACGCCAAGC  
 TACAACACAC CTTAAACACTC GCCTATTGTT AAGTGTGTC CTTGTGCGAT ACTGGTACTA ATGCGGTTCCG  
 BssHII  
 4061 GCGCAATTAA CCTCACTAA AGGGAACAA AGCTGGAAT CCACAATGAA CAATAATAAG ATTAATAATAG  
 CGCGTTAATT GGGAGTGAT TCCCTTGTTT TCGACCTTAA GGTGTACTT GTTATTATTC TAATTTTATC  
 4131 CTTGCCCCCG TTGCAGCGAT GGGTATTTTT TCTAGTAAA TAAAAGATAA ACTTAGACTC AAAACATTTA  
 GAACGGGGC AACGTGCTA CCCATAAAA AGATCATTTT ATTTCTATT TGAATCTGAG TTTTGTAAAT  
 4201 CAAAACAAC CCTAAAGTC CTAAAGCCCA AAGTGTATG CACGATCCAT AGCAAGCCCA GCCCAACCCA  
 GTTTTGTGTG GGGATTTCAG GATTTCGGGT TTCAAGATAC GTGCTAGGTA TCGTTCGGGT CCGGTTGGGT

FIG. 43G

4271 ACCCAACCCA ACCCAACCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC CCGGCACTAT CACCGTAGT  
 TGGGTGGGT TGGGTGGGT CACGTCGGT CACCGTTAT CAGAGGTGG GGCCTGATA GTGGCACTCA  
 4341 TGTCGGACC ACCGCAGTC TCGCAGCCAA AAAAAAATA AGAAGAAAA AAAAGAAAA GAAAAACAGC  
 ACAGGCGTG TGGCGTGCAG AGCGTCGGT TTTTTTTTTT TCTTCTTTT TTTTCTTTT CTTTTTCTCG  
 4411 AGGTGGTCC GGTCTGTGG GGC CGGAAA GCGAGGAGGA TCGCAGCAG CAGCAGGCC CCGCCCTCCC  
 TCCACCCAG CCCAGCACCC CCGGCTTTT CGCTCTTCT AGCGCTCGTC GCTGCTCCGG GCCGGGAGGG  
 4481 TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATAATACA TACCCCTCCC TCTCTCCA TCCCCCAAC  
 AGCGAAGGT TTCTTTCGG GGGTAGCGG TGATAATAT ATGGGGGGG AGAGGAGGT AGGGGGTTG  
 4551 CCTACACCA CCACCAACC CACCTCTCC CCCCTCGCT CCGGACGAC AGCTCTCTCC CCTCCCCCT  
 GGATGGTGGT GGTGGTGGT GTGGAGGAG GGGGAGCGAC GGCCTGCTGC TCGAGGAGGG GGGAGGGGA  
 4621 CCGCCGCCG CCGTAACCAC CCGCCCTCT TCTCTTCT TCTCTCTCT TTTTTTCTG CTCGCTCTCG  
 GCGCGCGCG GCCATTGGT GCGCGGGAG AGGAGAAAG AAGAGGCAA AAAAAAGCA GAGCCAGAGC  
 4691 ATCTTTGGC TTGGTAGTT GGTGGGCGA GAGCGCTTC GTGCCCCAGA TCGGTGCGG GAGGGGCGG  
 TAGAAACCG AACCATCAA CCCACCGCT CTCGCCGAG CAGCGGTCT AGCCACGCG CCTCCCCGCC  
 BamHI BglII  
 4761 GATCTCGCG CTGGCGTCT CCGCGGTGAG TCGGCCCGA TCCTCGCGG GAATGGGCT CTCGGATGTA  
 CTAGAGCGC GACCGCAGG GCCCGCACT AGCCGGGCT AGGAGCGCC CTTACCCCGA GAGCCTACAT  
 BglII  
 4831 GATCTCTTT CTCTCTCTT TTGTGTGTAG AATTGAATC CTTACGATTT GTTCATCGGT AGTTTTCTT  
 CTAGAGAAA GAAAGAAA AACACCATC TTAACCTTAG GGAGTCGTAA CAAGTAGCCA TCAAAAAAGAA  
 4901 TTCTATGATTT GTGACAAATG CAGCCTCGT CCGAGCTTT TTGTAGGTAG  
 AAGTACTAAA CACTGTTTAC GTCGAGCAC GCCTCGAAA AACATCCATC

FIG. 43H

**FIG. 44A**

ECORI  
1 AATCCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTTAGT  
TTAAGGTGTT ACTTGTTATT ATCTTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAAGATCA

71 AAAATAAAG ATAAACTTAG ACTCAAAA CAACCCCTAA AGTCTTAAAG CCCAAAGTGC  
TTTTATTTC TATTGAATC TGAGTTTGT AAATGTTTT GTTGGGATT TCAGGATTTC GGGTTTCACG

141 TATGCACGAT CCATAGCAAG CCCAGCCCA CCCAACCCAC CCCAGTCAG CCAACTGGCA  
ATACGTGCTA GGTATCGTTC GGTGCGGTT GGGTTGGGTT GGGTTGGGTG GGGTCACGTC GGTGACCGT

211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCCG CACCAACGCA CGTCTGCGAG CCAAAAAAAA  
TTATCAGAGG TGGGGCCGT GATAGTGGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTTTT

281 AAAAAGAAAG AAAAAAAGA AAAAAGAAA CAGCAGGTGG GTCCGGGTGG TGGGGGCCGG AAAAGCGAGG  
TTTTTCTTTC TTTTTTTTCT TTTTCTTTTT GTCTGTCACC CAGGCCCAGC ACCCCCGGCC TTTTCGCTCC

351 AGGATCGCGA GCAGCGACGA GGCCCGGCCC TCCCTCCGCT TCCTAAAGAA CGCCCCCAT CGCCACTATA  
TCCTAGCGCT CGTCGCTGCT CCGGGCCGGG AGGAGGCGA AGGTTCTTTT GCGGGGGGTA GCGGTGATAT

421 TACATACCC CCCCTCTCCT CCCATCCCC CAACCTACC ACCACACCA CCACCACTC CTCCCCCTC  
ATGTATGGG GGGAGAGGA GGTAGGGG GTTGGGATGG TGGTGGTGGT GGTGGTGGAG GAGGGGGGAG

491 GCTGCCGGAC GACGAGCTCC TCCCCCTCC CCTCCGCGG CCGCCGGTAA CCACCCCGCC CCTCTCCTCT  
CGACGGCCTG CTGCTCGAG AGGGGGGAGG GGGAGGCGGC GCGGCCCATT GGTGGGGCGG GAGAGAGGAGA

561 TTTCTTCTCC GTTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG  
AAGAAAGAGG CAAAAAAGA AGCAGAGCCA GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC

631 CTTGCTGCC CAGATCGGTG GCGGGAGGG GCGGGATCTC GCGCTGGCG TCTCCGGGCG TGAGTCGGCC  
GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC GCGCCTAGAG CCGCCGACCGC AGAGGCCCGC ACTCAGCCCG

BamHI  
701 CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGATAGATCTT CTTTCTTCTT TCTTTTGTG GTAGAATTTG  
GCCTAGGAGC GCCCCTTACC CCGAGAGCCT ACATCTAGAA GAAAGAAAGA AGAAAAACAC CATCTTAAAC

BglII  
TTTTTCTTCTT

FIG.\_44B

```

771  AATCCCTCAG CATTGTTTCAT CGGTAGTTTT TCTTTTCATG ATTTGTGACA AATGCAGCCT CGTCCGGAGC
    TTAGGGAGTC GTAAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTCG

    HindIII
    ~~~~~
841 TTTTTTTGTAG GTAGAAGCTT ACMATGCMCG TGCACAAGGA GGTSAACTTC GTSGCCTACC TCCTGATCGT
 AAAAAACATC CATCTTCGAA TGTACCKGC ACGTGTCCT CCASITTGAAG CASC GGATGG AGGACTAGCA

 NcoI
    ~~~~~
911  SCTCGGCTC CTCTTGCTCG TSTCCGCCAT GGAGCAGTG GACGCCAAGG CCTGCACCCK CGAGTGCAGC
    BGAGCCGGAG GAGAACGAGC ASAGGCGGTA CTTGCGCAC CTGCGGTTCC GGACGTGGGM GTCACGCCCG

    NotI
    ~~~~~
981 AACCTCGGCT TCGGCATCTG CCCGGCGGCC GCCTCCAGC AGGCATCTC CGAAGACCTC TACAGCCGTT
 TTGGAGCCGA AGCCGTAGAC GGGCCGCCGG CGGAGGTGG TCCCGTAGAG GCTTCTGGAG ATGTCGGCAA

 SalI
    ~~~~~
    AcoI
    ~~~~~
1051 TAGTCGAAAT GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA
 ATCAGCTTTA CCGGTGATAG AGGTTTCGAC GGATGCGGT GGACACGTTG TAAGGCAGCT GATAATAGTT

 BamHI
    ~~~~~
1121 GGGAGAGAAA ATTTACAATT CTCAAACTGA CATTAAAGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA
    CCCCTCTCTT TAAATGTTAA GAGTTTGACT GTAATTGCTT ACCTAGGAGG CGCTGCTGTC GTCGTTCTTT

1191 ATAATCACCG TCTTCCGTGG CACTGGTAGT GATACGAATC TACRACTCGA TACTAACTAC ACCCTCACGC
    TATTAGTGGC AGAAGGCACC GTGACCATCA CTATGCTTAG ATGTTGAGCT ATGATTGATG TGGGAGTGGC

1261 CTTTCCGACAC CCTACCACAA TGCACCGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT
    GAAAGCTGTG GGATGGTGT ACCTTGCCAA CACTTCATGT GCCACCTATA ATATAACCTA CCCAGAGGCA

```

**FIG.\_44C**

1331 CCAGGACCAA GTCGAGTCGC TTGTCAAACA GCAGGTTAGC CAGTATCCGG ACTACGGCT GACCGTGACC  
 GGTCCCTGGTT CAGCTCAGCG AACAGTTTGT CGTCCAATCG GTCATAGGCC TGATGCGCGA CTGGCACTGG  
 1401 GGCCACKCCC TCGGCGCCTC CCTGGGGCA CTCACTGCCG CCCAGCTGTC TCGACATAC GACAAACATCC  
 CCGGTGCGGG AGCCGCGGAG GGACCGCCGT GAGTGACGGC GGGTCGACAG ACGCTGTATG CTGTTGTAGG  
 1471 GCCTGTACAC CTTTCGGCGAA CCGCGCAGCG GCAATCAGGC CTTCCGCTCG TACATGAACG ATGCCCTTCCA  
 CGGACATGTG GAAGCCGCTT GCGCGCTGCG CGTTAGTCCG GAAGCGCAGC ATGTACTTGC TACGGAAGGT  
 XhoI  
 ~~~~~  
 1541 AGCCTCGAGC CCAGATACGA CGCAGTATTT CCGGGTCACT CATGCCAACG ACGGCATCCC AAACCTGCCC
 TCGGAGCTCG GGTCTATGCT GCGTCATAAA GCGCCAGTGA GTACGGTTGC TGCCGTAGGG TTTGGACGGG
 NcoI
 ~~~~~  
 1611 CCGGTGGAGC AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAAC  
 GGCCACCCTCG TCCCCATGCG GGTACCGCCA CATCTCATGA CCTCGCACT AGGAATGTCTG CCGGTCTTGT  
 1681 CATTTGTCTG CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGGCGGA CAGGGTGTGA ATAATGCCGA  
 GTAAACAGAC GTGACCCCTTA CTTACAGTCA CGACACTCCG GGTCCCGCT GTCCACACT TATTACGCGT  
 SphI  
 ~~~~~  
 1751 CACGACTTAT TTTGGGATGA CGAGCGGCGC ATGCACCTGG CCGGTGCGGG CCGCGGAAC CACTGAAGGA
 GTGCTGAATA AAACCTACTT GCTCGCGCGG TACGTGGACC GGCCAGCGCC GGCCTTCTG GTGACTTCTT
 1821 TGAGCTGTAA AGAAGCAGAT CGTTCAAACA TTTGGCAATA AAGTTTCTTA AGATTGAATC CTGTTGCCGG
 ACTCGACATT TCTTCGTCTA GCAAGTTTGT AAACCGTTAT TTCAAAGAAT TCTAACTTAG GACAACGGCC
 1891 TCTTGGGATG ATTATCATAT AATTTCGTGT GAATTACGTT AAGCATGTAA TAATTAACAT GTAATGCATG
 AGAACGCTAC TAATAGTATA TTAAGACAA CTTAATGCAA TTCGTACATT ATTAATTGTA CATACGTAC
 1961 ACGTTATTTA TGAGATGGGT TTTTATGATT AGAGTCCCGC AATTATACAT TTAATACGCG ATAGAAAAA
 TGCAATAAAT ACTCTACCCA AAAATACTAA TCTCAGGGCG TTAATATGTA AATTATGCGC TATCTTTTGT

FIG.-4D

| | BssHII
~~~~~ | BssHII
~~~~~ | ClaI HindIII
~~~~~ | XbaI
~~ |
|------|---|-----------------|-----------------------|------------|
| 2031 | AAATATAGCG CGCAAACTAG GATAAATTAT CGCGCGCGGT GTCATCTATG TTACTAGATC GATAAGCTTC
TTTATATCGC GCGTTTGATC CTATTTAATA GCGCGCGCCA CAGTAGATAC AATGATCTAG CTATTGGAAG | | | |
| | XbaI
~~~~~ | BssHII
~~~~~ | | |
| 2101 | TAGAGCGGCC GGTGGAGCTC CAATTGCGCC TATAGTGAGT CGTATTACGC GCGCTCACAG GCCGTGCTTT
ATCTCGCCGG CCACCTCGAG GTTAAGCGGG ATATCACTCA GCATAATGCG CGCGAGTGAC CGGCAGCAAA | | | |
| 2171 | TACAACGTCTG TGACTGGGAA AACCTTGGCG TTACCCAACT TAATCGCCTT GCAGCACATC CCCCTTTTCG
ATGTTGCAGC ACTGACCCTT TTGGGACCGC AATGGGTTGA ATTAGCGGAA CGTCGTGTAG GGGGAAAGCG | | | |
| 2241 | CAGCTGGCGT AATAGCGAAG AGGCCCGCAC CGATCGCCCT TCCCAACAGT TCGGCAGCCT GAATGGCGAA
GTCGACCGCA TTATCGCTTC TCCGGGCGTG GCTAGCGGGA AGGGTTGTCA ACGCGTCGGA CTTACCGCTT | | | |
| 2311 | TGGGACGCGC CCTGTAGCGG CGCATTAAGC GCGGCGGGTG TGGTGGTTAC GCGCAGCGTG ACCGCTACAC
ACCTGCGCG GGAATCGCC GCGTAATTCT GCGCGCCAC ACCACCAATG CGCGTCGCAC TGGCGATGTG | | | |
| 2381 | TTGCCAGCGC CCTAGCGCCC GCTCCCTTTC GCTTCTTCCC CTTCCTTTC GGCACGTTTC CCGCTTTTCC
AACGGTCCG GGAATCGGG CGAGGAAAGC GAAAGAGGG AAGGAAAGAG CGGTGCAAGC GGCAGAAAGG | | | |
| 2451 | CCGTCAAGCT CTAAATCGGG GGCTCCCTTT AGGTTCCGA TTTAGTGCTT TACGGCACCT CGACCCCAAA
GGCAGTTCTGA GATTAGCCC CCGAGGAAA TCCCAAGGCT AATCACGAA ATGCCGTGGA GCTGGGGTTT | | | |
| 2521 | AAACTTGATT AGGGTGATGG TTCACGTAGT GGGGCATCGC CCTGATAGAC GGTTTTTCGC CCTTTGACGT
TTTGAACATA TCCCACTACC AAGTGCAATCA CCGGATAGCG GGAATATCTG CCAAAAAGCG GGAATACTGA | | | |
| 2591 | TGGAGTCCAC GTTCTTTAAT AGTGGACTCT TGTTCCAAAC TGAACAACA CTCAACCCCTA TCCTCGGTCTA
ACCTCAGGTG CAAGAAATTA TCACCTGAGA ACAAGGTTG ACCTTGTTGT GAGTTGGGAT AGAGCCAGAT | | | |
| 2661 | TTCTTTTGAT TTATAAGGGA TTTTGGCGAT TTCGGCTAT TGGTTAAAAA ATGAGCTGAT TTAACAAAAA
AAGAAAATA AATATCCCT AAAACGGCTA AAGCGGATA ACCAATTTT TACTCGACTA AATGTGTTTT | | | |

FIG._44E

2731 TTTAACGGCA ATTTTAAACA AATATTAAACG CTTACAATTT AGTGGCACT TTTTCGGGGA ATGTGCGCGG
AAATGCGCT TAAAATTGTT TTATAATTGC GAATGTTAAA TCACACCGTGA AAAGCCCCCTT TACACGCGCC

2801 AACCCCTATT TGTATTATTT TCTAAATACA TTCAAAATATG TATCCGCTCA TGAGACAATA ACCCTGATAA
TTGGGGATAA ACAATAAAA AGATTATGT AAGTTATATC ATAGGCGAGT ACTCTGTTAT TGGGACTATT

2871 ATGCTTCAAT AATATTGAAA AAGGAAGAGT ATGAGTATTC AACATTTCG TGTCGCCCTT ATTCCCCTTT
TACGAAGTTA TTATAACTTT TTCTTCTCA TACTCATAG TTGTAAAGGC ACAGCGGGA TAAGGGAAAA

2941 TTGCGGCATT TTGCTTCTT TTTTTCCTC ACCCAGAAAC GCTGGTGAAA GTAAAAGATG CTGAAGATCA
AACGCCGTA AACGGAAGGA CAAAACGAG TGGGTCTTTG CGACCACCTT CATTTCTAC GACTTCTAGT

3011 GTTGGGACA CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAAG TCCTTGAGAG TTTTCGCCCC
CAACCCAGT GCTCACCCAA TGTAGCTGA CCTAGATTG TCGCCATTCT AGGAACCTCTC AAAAGCGGGG

3081 GAAGAACGTT TTCCAATGAT GAGCATTCTT AAAGTTCTGC TATGTGGCGC GGTATTATCC CGTATTGACG
CTTCTTGCAA AAGGTTACTA CTCGTGAAA TTTCAAGACG ATACACCGCG CCATAATAGG GCATAACTGC

3151 CCGGGCAAGA GCAACTCGGT CGCCGCATAC ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC
GGCCCGTTCT CGTTGAGCCA GCGCGGTATG TGATAAGAGT CTTACTGAAC CAACTCATGA GTGGTCAGTG

3221 AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGATTTA TGCAGTCTG CCATAACCAT GAGTGATTAAC
TCCTTTTCGTA GAATGCCTAC CGTACTGTCA TTCTCTTAAT ACGTCACGAC GGTATTGTA CTCACTATTG

3291 ACTGCGCCA ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC CGCTTTTTTG CACAACATGG
TGACGCCGGT TGAATGAAGA CTGTTGCTAG CCTCCTGGCT TCCTCGATTG GCGAAAAAAC GTGTTGTACC

3361 GGGATCATGT AACTCGCTT GATCGTTGGG AACCGAGCT GAATGAAGCC ATACCAAACG ACGAGCGTGA
CCCTAGTACA TTGAGCGGAA CTAGCAACCC TTGGCCTCGA CTTACTTCGG TATGGTTTGC TGCTCGCACT

3431 CACCACGATG CCTGTAGCAA TGGCAACAAC GTTGGCGCAA CTATTAACTG GCGAACTACT TACTCTAGCT
GTGTGCTAC GGACATCGTT ACCGTTGTTG CAACGCGTTT GATAATTGAC CGCTTGATGA ATGAGATCGA

3501 TCCCGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG TTGCAGGACC ACTTCTGCGC TCGGCCCTTC
AGGGCCGTTG TTAATTATCT GACCTACCTC CGCCTATTTT AACGTCCTGG TGAAGACGCG AGCCGGBAG

FIG._44F

3571 CGGCTGGCTG GTTATTTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT CGCGGTATCA TTGCAGCACT
 GCCGACCGAC CAAATAACGA CTATTTAGAC CTCGGCCACT CGCACCCAGA GCGCCATAGT AACGTCGTGA
 3641 GGGGCCAGAT GGTAGCCCT CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC TATGGATGAA
 CCCCCTCTA CCAATTCGGG GGGCATAGCA TCAATAGATG TGCTGCCCTT CAGTCCGTG ATACCTACTT
 3711 CGAAATAGAC AGATCCCTGA GATAGGTGCC TCACTGATTA AGCAATGGTA ACTGTCAGAC CAAGTTTACT
 GCITTTATCTG TCTAGCGACT CTATCCACGG AGTGACTAAT TCGTAACCAT TGACAGTCTG GTTCAAAATGA
 3781 CATATATACT TTAGATTGAT TTAATACTTC ATTTTAAAT TAAAAGGATC TAGGTGAAGA TCCTTTTGA
 GTATATATGA AATCTAACTA AATTTTGAAG TAAAAATTAA ATTTCCCTAG ATCCACTTCT AGGAAAAACT
 3851 TAATCTCATG ACCAAAATCC CTTAACGTGA GTTTTTCGTT CACTGAGCGT CAGACCCCGT AGAAAAGATC
 ATTAGAGTAC TGGTTTTAGG GAATGCACT CAAAAGCAAG GTGACTCGCA GTCTGGGCA TCTTTTCTAG
 3921 AAAGGATCTT CTTGAGATCC TTTTTCCTG CGGTAATCT GCTGCTTGCA AACAAAAA CCACCGCTAC
 TTTCCCTAGAA GAACTCTAGG AAAAAAGAC GCGCATTAGA CGACGAACGT TTGTTTTTTT GGTGGCGATG
 3991 CAGCGGTGCT TTGTTTGCCG GATCAAGAGC TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC
 GTCGCCACCA AACAAACGGC CTAGTTCTCG ATGGTTGAGA AAAAGGCTTC CATTGACCGA AGTCGTCTCG
 4061 GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA GGCACCACT TCAAGAATC TGTAGCACCG
 CGTCTATGGT TTATGACAGG AAGATCACAT CGGCATCAAT CCGGTGGTGA AGTTCTTGAG ACATCGTGCC
 4131 CCTACATACC TCGTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG CGATAAGTCC TGTCTTACC
 GGATGTATGG AGCGAGACGA TTAGGACAAT GGTACCCGAC GACGGTCACC GCTATTGAGC ACAGAAATGGC
 4201 GGTGGACTC AAGACGATAG TTACCGGATA AGGCGAGCG GTCGGGCTGA ACGGGGGTT CGTGACACAA
 CCAACCTGAG TTCTGCTATC AATGGCCTAT TCCGCGTCCG CAGCCCGACT TGCCCCCAA GCACGTGTGT
 4271 GCCAGCTTG GAGCGAACGA CCTACACCGA ACTGAGATAC CTACAGCGTG AGCTATGAGA AAGCGCCACG
 CCGGTCCGAC CTCGCTTGCT GGATGTGGCT TGACTCTATG GATGTCGCAC TCGATACTCT TTCGCGGTGC
 4341 CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG GCAGGGTCCG AACAGGAGAG CGCACAGAGG
 GAAGGGCTTC CCTCTTTCCG CCTGTCCATA GGCCATTCGC CGTCCAGCC TTGTCTCTC GCGTGCTCCC

FIG. 44G

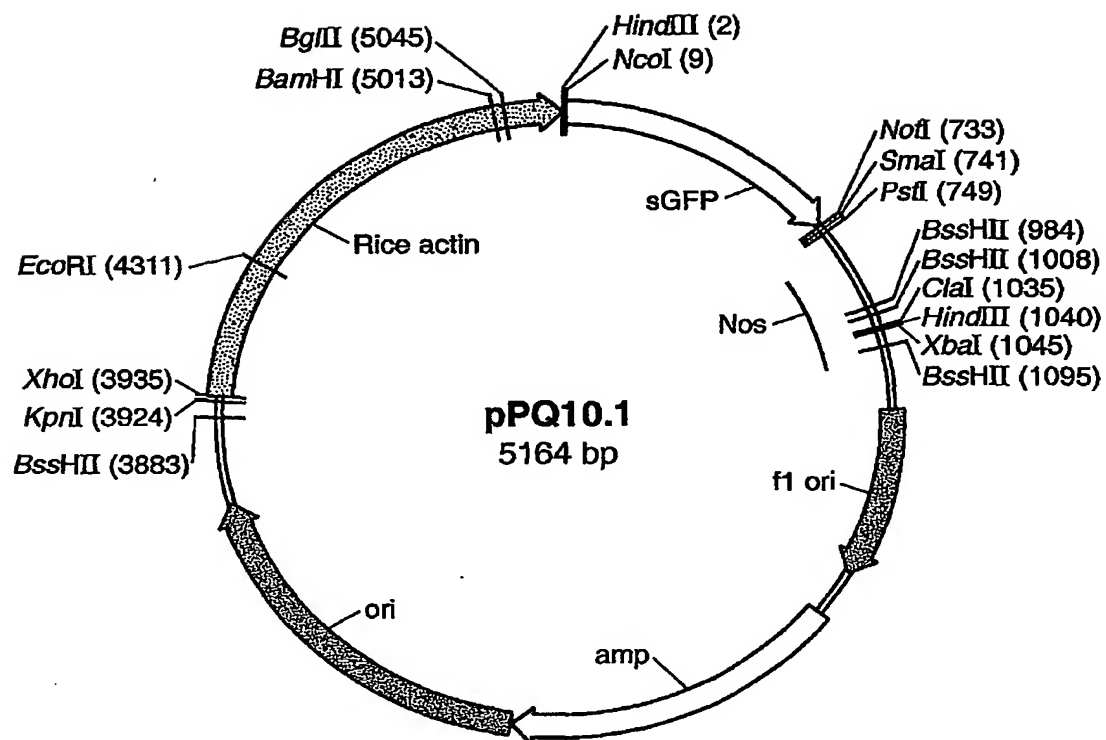
4411 AGCTTCCAGG GGGAAACGCC TGGTATCTTT ATAGTCTGT CGGTTTTCGC CACCTCTGAC TTGAGCGTTCG
 TCGAAGGTCC CCTTTGCGG ACCATAGAAA TATCAGGACA GCCCAAAGCG GTGGAGACTG AACTCGCAGC
 4481 ATTTTGTGA TGCTCGTCAG GGGGGCGGAG CCTATGGA AAACCCAGCA ACGCGSCCTT TTTACGGTTC
 TAAAAACACT ACGAGCAGTC CCCCCGCCTC GGATACCTTT TTGCGGTCTGT TCGCGCGGAA AAATGCCAAG
 4551 CTGGCCCTTT GCTGGCCCTT TGCTCACATG TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAAACCGTA
 GACCGGAAAA CGACCGGAAA ACGAGTGTAC AAGAAAGGAC GCAATAGGGG ACTAAGACAC CTATTGGCAT
 4621 TTACCGCCCTT TGAGTGAGCT GATACCGCTC GCCGACGCCG AACGACCGAG CGCAGCGAGT CAGTGAGCGA
 AATGGCGGAA ACTCACTCGA CTATGGCGAG CGGCGTCGGC TTGCTGGCTC GCGTCGCTCA GTCACTCGCT
 4691 GGAAGCGGAA GAGCGCCCAA TACGCAAAACC GCTCTTCCCC GCGCGTTGGC CGATTCAATTA ATGCAGCTGG
 CCTTCGCCCTT CTCGCGGGTT ATGCGTTTGG CGGAGAGGGG CGCGCAACCG GCTAAGTAAT TACGTCGACC
 4761 CACGACAGGT TTCCCGACTG GAAAGCGGGC AGTGAGCGCA ACGCAATTA TGTGAGTTAG CTCACCTCAT
 GTGCTGTCCA AAGGGCTGAC CTTTCGCCCG TCACTCGCGT TCGGTAAAT ACACTCATC GAGTGAGTAA
 4831 AGGCACCCCA GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAAT
 TCCGTGGGGT CCGAAATGTG AAATACGAAG GCCGAGCATA CAACACACCT TAACACTCGC CTATTGTATA

 BssHII
 ~~~~~  
 4901 TCACACAGGA AACAGCTATG ACCATGATTA CGCCAAGCGC GCAATTAACC CTCACTAAAG GGAACAAAAG  
 AGTGTGTCCT TTGTCGATAC TGGTACTAAT GCGGTTCGGG CGTTAATTGG GAGTGATTTT CCTTGTTC

EcoR

4971 CTGG  
 GACC

FIG. 44H

**FIG.\_45A**

## Sequence for pPQ10.1

HindIII NcoI  
-----

1 AAGCTTACCA TGGTGAGCAA GGGCGAGGAG CTGTTACCG GGGTGGTGCC  
CATCCTGGTC GAGCTGGACG  
TTCGAATGGT ACCACTCGTT CCCGCTCCTC GACAAGTGGC CCCACCACGG  
GTAGGACCAG CTCGACCTGC

71 GCGACGTGAA CGGCCACAAG TTCAGCGTGT CCGGCGAGGG CGAGGGCGAT  
GCCACCTACG GCAAGCTGAC  
CGCTGCACTT GCCGGTGTTC AAGTCGCACA GGCCGCTCCC GCTCCCGCTA  
CGGTGGATGC CGTTCGACTG

141 CCTGAAGTTC ATCTGCACCA CCGGCAAGCT GCCCGTGCCC TGGCCCACCC  
TCGTGACCAC CTTACCTTAC  
GGACTTCAAG TAGACGTGGT GGCCTTCGA CGGGCACGGG ACCGGGTGGG  
AGCACTGGTG GAAGTGGATG

211 GCGGTGCAGT GCTTCAGCCG CTACCCCGAC CACATGAAGC AGCACGACTT  
CTTCAAGTCC GCCATGCCCG  
CCGCACGTCA CGAAGTCGGC GATGGGGCTG GTGTACTTCG TCGTGCTGAA  
GAAGTTCAGG CGGTACGGGC

281 AAGGCTACGT CCAGGAGCGC ACCATCTTCT TCAAGGACGA CGGCAACTAC  
AAGACCCGCG CCGAGGTGAA  
TTCCGATGCA GGTCTCTCGG TGGTAGAAGA AGTTCCTGCT GCCGTTGATG  
TTCTGGGCGC GGCTCCACTT

351 GTTCGAGGGC GACACCCTGG TGAACCGCAT CGAGCTGAAG GGCATCGACT  
TCAAGGAGGA CGGCAACATC  
CAAGCTCCCG CTGTGGGACC ACTTGGCGTA GCTCGACTTC CCGTAGCTGA  
AGTTCCTCCT GCCGTTGTAG

421 CTGGGGCACA AGCTGGAGTA CAACTACAAC AGCCACAACG TCTATATCAT  
GGCCGACAAG CAGAAGAACG  
GACCCCGTGT TCGACCTCAT GTTGATGTTG TCGGTGTTGC AGATATAGTA  
CCGGCTGTTT GTCTTCTTGC

491 GCATCAAGGT GAACTTCAAG ATCCGCCACA ACATCGAGGA CGGCAGCGTG  
CAGCTCGCCG ACCACTACCA  
CGTAGTTCCA CTTGAAGTTC TAGGCGGTGT TGTAGCTCCT GCCGTGCGAC  
GTCGAGCGGC TGGTGATGGT

561 GCAGAACACC CCCATCGGCG ACGGCCCCGT GCTGCTGCCC GACAACCACT  
ACCTGAGCAC CCACTCCGCC  
CGTCTGTGG GGGTAGCCGC TGCCGGGGCA CGACGACGGG CTGTTGGTGA  
TGGACTCGTG GGTACGGCGG

631 CTGAGCAAAG ACCCCAACGA GAAGCGCGAT CACATGGTCC TGCTGGAGTT  
CGTGACCGCC GCCTGGATCA

Fig. 45B

CCGCATTATC GCTTCTCCGG GCGTGGCTAG CGGGAAGGGT TGTCAACGCG  
TCGGACTTAC CGCTTACCCCT

1261 CGCGCCCTGT AGCGGCGCAT TAAGCGCGGC GGGTGTGGTG GTTACGCGCA  
GCGTGACCGC TACACTTGCC  
GCGCGGGACA TCGCCGCGTA ATTGCGCGCG CCCACACCAC CAATGCGCGT  
CGCACTGGCG ATGTGAACGG

1331 AGCGCCCTAG CGCCCGCTCC TTTCGCTTTC TTCCCTTCCT TTCTCGCCAC  
GTTGCGCCGC TTTCCCGTC  
TCGCGGGATC GCGGGCGAGG AAAGCGAAG AAGGGAAGGA AAGAGCGGTG  
CAAGCGGCCG AAAGGGGCAG

1401 AAGCTCTAAA TCGGGGGCTC CCTTAGGGT TCCGATTAG TGCTTTACGG  
CACCTCGACC CCAAAAACT  
TTCGAGATT AGCCCCGAG GGAAATCCCA AGGCTAAATC ACGAAATGCC  
GTGGAGCTGG GGTTTTTTGA

1471 TGATTAGGGT GATGGTTCAC GTAGTGGGCC ATCGCCCTGA TAGACGGTTT  
TTCGCCCTTT GACGTTGGAG  
ACTAATCCCA CTACCAAGTG CATCACCCGG TAGCGGGACT ATCTGCCAAA  
AAGCGGGAAA CTGCAACCTC

1541 TCCACGTTCT TTAATAGTGG ACTCTTGTT CAACTGGAA CAACACTCAA  
CCCTATCTCG GTCTATTCTT  
AGGTGCAAGA AATTATCACC TGAGAACAAG GTTTGACCTT GTTGTGAGTT  
GGGATAGAGC CAGATAAGAA

1611 TTGATTATA AGGGATTTTG CCGATTTGCG CCTATTGGTT AAAAAATGAG  
CTGATTAAAC AAAAATTTAA  
AACTAAATAT TCCCTAAAC GGCTAAAGCC GGATAACCA TTTTTTACTC  
GACTAAATTG TTTTAAATT

1681 CGCGAATTTT AACAAAATAT TAACGCTTAC AATTTAGGTG GCACTTTTCG  
GGGAAATGTG CGCGGAACCC  
GCGCTTAAAA TTGTTTTATA ATTGCGAATG TTAAATCCAC CGTGAAAAGC  
CCCTTTACAC GCGCCTTGGG

1751 CTATTTGTTT ATTTTCTAA ATACATTCAA ATATGTATCC GCTCATGAGA  
CAATAACCCCT GATAAATGCT  
GATAAACAAA TAABAAGATT TATGTAAGTT TATACATAGG CGAGTACTCT  
GTTATTGGGA CTATTTACGA

1821 TCAATAATAT TGAAAAAGGA AGAGTATGAG TATTCAACAT TTCCGTGTCG  
CCCTTATTCC CTTTTTTCG  
AGTTATTATA ACTTTTTCCT TCTCATCTC ATAAGTTGTA AAGGCACAGC  
GGGAATAAGG GAAAAAACGC

1891 GCATTTTGCC TTCTGTTTT TGCTACCCA GAAACGCTGG TGAAAGTAA  
AGATGCTGAA GATCAGTTGG  
CGTAAACGG AAGGACAAA ACGAGTGGGT CTTTGCGACC ACTTTCATTT  
TCTACGACTT CTAGTCAACC

Fig. 45C

1961 GTGCACGAGT GGGTTACATC GAACTGGATC TCAACAGCGG TAAGATCCTT  
GAGAGTTTTC GCCCGAAGA  
CACGTGCTCA CCCAATGTAG CTTGACCTAG AGTTGTCGCC ATTCTAGGAA  
CTCTCAAAAG CGGGGCTTCT

2031 ACGTTTTCCA ATGATGAGCA CTTTTAAAGT TCTGCTATGT GCGCGGGTAT  
TATCCCGTAT TGACGCCGGG  
TGCAAAAGGT TACTACTCGT GAAAATTTCA AGACGATACA CCGCGCCATA  
ATAGGGCATA ACTGCGGCCC

2101 CAAGAGCAAC TCGGTCGCCG CATACACTAT TCTCAGAATG ACTTG GTTGA  
GTACTCACCA GTCACAGAAA  
GTTCTCGTTG AGCCAGCGGC GTATGTGATA AGAGTCTTAC TGAACCAACT  
CATGAGTGGT CAGTGTCTTT

2171 AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA  
ACCATGAGTG ATAACACTGC  
TCGTAGAATG CCTACCGTAC TGTCATTCTC TTAATACGTC ACGACGGTAT  
TGGTACTCAC TATTGTGACG

2241 GGCCAACCTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT  
TTTTGCACAA CATGGGGGAT  
CCGTTGAAT GAAGACTGTT GCTAGCCTCC TGGCTTCCTC GATTGGCGAA  
AAAACGTGTT GTACCCCTA

2311 CATGTAACCT GCCTTGATCG TTGGGAACCG GAGCTGAATG AAGCCATACC  
AAACGACGAG CGTGACACCA  
GTACATTGAG CGGAAGTAGC AACCCTTGGC CTCGACTTAC TTCGGTATGG  
TTTGCTGCTC GCACTGTGGT

2381 CGATGCCTGT AGCAATGGCA ACAACGTTGC GCAAACTATT AACTGGCGAA  
CTACTTACTC TAGCTTCCCG  
GCTACGGACA TCGTTACCGT TGTTGCAACG CGTTTGATAA TTGACCGCTT  
GATGAATGAG ATCGAAGGGC

2451 GCAACAATTA ATAGACTGGA TGGAGGCGGA TAAAGTTGCA GGACCACTTC  
TGCGCTCGGC CCTTCCGGCT  
CGTTGTTAAT TATCTGACCT ACCTCCGCCT ATTTCAACGT CTTGGTGAAG  
ACCGGAGCCG GGAAGGCCGA

2521 GGCTGGTTTA TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG  
TATCATTGCA GCACTGGGGC  
CCGACCAAAT AACGACTATT TAGACCTCGG CCACTCGCAC CCAGAGCGCC  
ATAGTAACGT CGTGACCCCG

2591 CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG  
GCAACTATGG ATGAACGAAA  
GTCTACCATT CGGGAGGGCA TAGCATCAAT AGATGTGCTG CCCCTCAGTC  
CGTTGATACC TACTTGCTTT

2661 TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACTGT  
CAGACCAAGT TTAATCATAT  
ATCTGTCTAG CGACTCTATC CACGGAGTGA CTAATTCGTA ACCATTGACA  
GTCTGGTTCA AATGAGTATA

Fig. 45D

```

2731 ATACTTTAGA TTGATTTAAA ACTTCATTTT TAATTTAAA GGATCTAGGT GAAGATCCTT TTTGATAATC
TATGAAATCT AACTAAATTT TGAAGTAAAA ATTAAATTTT CCTAGATCCA CTTCCTAGGAA AAACATATTAG

2801 TCATGACCAA AATCCCTTAA CGTGACTTTT CGTTCCACTG AGGTCAGAC CCGTAGAAA AGATCAAAGG
AGTACTGGTT TTAGGGAATT GCACCTCAAAA GCAAGGTGAC TCGCAGTCTG GGGCATCTTT TCTAGTTTCC

2871 ATCTTCTTGA GATCCTTTTT TTCTGGCGGT AATCTGCTGC TTGCAAAACAA AAAAACACC GCTACCAGCG
TAGAAGAACT CTAGGAAAAA AAGACGGCA TTAGACGACG AACGTTTGT TTTTGGTGG CGATGGTCCG

2941 GTGGTTTGT TTCCCGGATCA AGACTACCA ACTCTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA
CACCAAAACAA ACGGCCTAGT TCTCGATGGT TGAGAAAAAG GCTTCCATTG ACCGAAGTCG TCTCGCGTCT

3011 TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACCTTCAAG AACTCTGTAG CACCGCCTAC
ATGGTTTATG ACAGGAAGAT CACATCGGCA TCAATCCGGT GGTGAAGTTC TTGAGACATC GTGGCGGATG

3081 ATACCTCGCT CTGCTAATCC TGTACCAGT GGTGCTGCC AGTGCGATA AGTCGTGTCT TACCGGGTTG
TATGGAGCGA GACGATTAGG ACAATGGTCA CCGACGACGG TCACCCTAT TCAGCACAGA ATGGCCCAAC

3151 GACTCAAGAC GATAGTTACC GGATAAGCG CAGCGGTGG GGTTCGTGC ACACAGCCCA
CTGAGTCTG CTATCAATGG CCTATTCCG GTCCGCCAGCC CGACTTGCCC CCCAAGCACG TGTGTGGGT

3221 GCTTGGAGCG AACGACCTAC ACCGAACTGA GATACCTACA GCGTGAGCTA TGAGAAAGCG CCACGCTTCC
CGAACCTCGC TTGCTGGATG TGGCTTGACT CTATGGATGT CGCACTCGAT ACTCTTTCGC GGTGCGAAGG

3291 CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT
GCTTCCCTCT TTCCCGCCTGT CCATAGGCCA TTCGCCGTCC CAGCCTTGT CTCTCGCGTG CTCCCCTCGAA

3361 CCAGGGGGAA ACGCTGGTA TCTTTATAGT CCTGTGCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT
GGTCCCCCTT TCGGACCAT AGAATATCA GGCAGGCCCA AAGCGGTGA GACTGAATC GCAGCTAAA

3431 TGTGATGCTC GTCAGGGGG CGGAGCCTAT GGA AAAACGC CAGCAACCG GCCTTTTAC GGTTCCTGGC
ACACTACGAG CAGTCCCCC GCCTCGGATA CCTTTTTCG GTCTGTGGC CGGAAAAATG CCAAGGACCG

```

**FIG. 45E**

```

3501 CTTTTGCTGG CTTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATTA CCGTATTACC
    GAAACGACC GGAACGAG TGTACAAGAA AGGACGCAAT AGGGACTAA GACACCTATT GGCATAATGG

3571 GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG AGCGAGGAG
    CGAAACATCA CTCGACTATG GCGAGCGGCG TCGGCTTGCT GGTCTGCGTC GCTCAGTCAC TCGCTCCTTC

3641 CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCCGCGG TGGCCCGATT CATTAATGCA GCTGGCACGA
    GCCTTCTCGC GGGTATGCG TTTGGCGGAG AGGGGCGGCG AACCGGCTAA GTAAATTACGT CGACCGTGT

3711 CAGGTTTCCC GACTGGAAG CGGGCAGTGA CGGCAACGCA ATTAATGTGA GTTAGCTCAC TCATTAGGCA
    GTCCAAAGGG CTGACCTTTC GCCCGTCACT CGCGTTGCGT TAATTACACT CAATCGAGTG AGTAAATCCGT

3781 CCCCAGGCTT TACACTTTAT GCTTCCGGCT CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTTTACA
    GGGGTCCGAA ATGTGAATA CGAAGGCCGA GCATACAACA CACCTTAACA CTCGCCATT GTTAAAGTGT

                                     BssHII
3851 CAGGAACACG CTATGACCAT GATTACGCCA AGCGGCGCAAT TAACCTTCAC TAAAGGGAAC AAGAGCTGGG
    GTCCTTTGTG GATACTGGTA CTAATGCGGT TCGCGCGTTA ATTGGGAGTG ATTTCCCTTG TTTTCGACCC

                                     KpnI
                                     XhoI
3921 TACCGGGCCC CCCCTCGAGG TCATTTCATAT GCCTTGAAG AGAGTCGGGA TAGTCCAAA TAAACAAAG
    ATGCCCCGGG GGGGAGCTCC AGTAAGTATA CGAACTCTTC TCTCAGCCCT ATCAGGTTTT ATTTTCTTTTC

3991 GTAAGATTAC CTGGTCAAAA GTGAAAACAT CAGTTAAAAG GTGGTATAAG TAAATATCG GTATATAAG
    CATTCTAATG GACCAGTTTT CACTTTTGTG GTCAATTTTC CACCATATTC ATTTTATAGC CATTAATTTTC

4061 GTGGCCCAAA GTGAATTTA CTCTTTTCTA CTATTAAAA AATTGAGGAT GTTTTGTGCG TACTTTTGATA
    CACCGGGTTT CACTTTAAAT GAGAAAAGAT GATAATATTT TTAACCTCCTA CAAACAGCC ATGAACATAT

```

FIG. 45F



```

4131 CGTCATTTTT GTATGAATTG GTTTTTAAGT TTATTCGCGA TTGGAAATG CATATCTGTA TTGAGTCGG
GCAGTAAAAA CATACTTAAC CAAAAATCA AATAAGCGCT AAACCTTTAC GTATAGACAT AAACTCAGCC

4201 TTTTAAAGTT CGTTGCTTTT GTAAATACAG AGGATTGTG ATAAGAAATA TCTTTAAAAA ACCCATATGC
AAAAATTCAA GCAACGAAAA CATTTATGTC TCCCTAAACA TATCTTTTAT AGAAATTTTT TGGGTATACG

      EORI
      *****
4271 TAATTGACA TAATTTTGA GAAAAATATA TATTCAGGG TATCCACAA TGAACAATAA TAAGATTAA
ATAAACTGT ATAAAAACT CTTTTATAT ATAAGTCCG TTAAGGTGTT ACTTGTATT ATTCTAATTT

4341 ATAGCTTGCC CCCGTTGCAG CGATGGGTAT TTTTCTTAGT AAAATAAAG ATAACTTAG ACTCAAAACA
TATCGAACGG GGGCAACGTC GCTACCCATA AAAAGATCA TTTTATTTTC TATTGAATC TGAGTTTTGT

4411 TTTACAAAAA CACCCCTAA AGTCCTAAG CCCAAAGTGC TATGCACGAT CCATAGCAAG CCCAGCCCCA
AAATGTTTTT GTTGGGANT TCAGGATTC GGGTTTCACG ATACGTGCTA GGTATCGTTC GGGTCGGGTT

4481 CCCAACCCAA CCCAACCCAC CCAGTGCAG CCAACTGGCA AATAGTCTCC ACCCCGGCA CTATCACCGT
GGGTTGGGTT GGGTTGGGTG GGGTCACGTC GGTGACCGT TTATCAGAGG TGGGGGCCGT GATAGTGGCA

4551 GAGTTGTCG CACCACGCA CGTCTCGCAG CCAAAAAA AAAAGAAAG AAAAAAAGA AAAAGAAAAA
CTCAACAGGC GTGTGGCGT GCAGAGCGTC GGTTTTTTTT TTTTCTTTTC TTTTCTTTCT TTTTCTTTT

4621 CAGCAGGTGG GTCCGGGTG TGGGGGCCGG AAAAGCGAG AGGATCGCGA GCAGCGACGA GGGCCGGCCC
GTCGTCCACC CAGGCCAGC ACCCCGGCC TTTTTCGCTCC TCTAGCGCT CGTCGCTGCT CCGGGCCGGG

4691 TCCCTCCGCT TCCAAAGAA CGCCCCCAT CGCCACTATA TACATACCCC CCCCTCTCTT CCCATCCCCC
AGGAGGGCGA AGGTTCTTTT GCGGGGGGTA GCGGTGATAT ATGTATGGGG GGGGAGAGGA GGGTAGGGGG

4761 CAACCTACC ACCACACCA CCACCACCTC CTCCCCCTC GTGCCGGAC GACGAGCTCC TCCCCCTCC
GTTGGGATGG TGGTGGTGGT GGTGTGGAG GAGGGGGAG CGACGGCCTG CTGCTCGAGG AGGGGGAGG

4831 CCTCCGCCG CCGCCGGTAA CCACCCGCC CCTCTCTCTT TCTTTCTCTC GTTTTTTTTT TCGTCTCGGT
GGGAGGCCGC GCGGCCATT GGTGGGCCG GGAGAGGAGA AAGAAAGAG CAAAAAAA AGCAGAGCCA

```

**FIG. 45G**

4901 CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG CTTCTGTCGCC CAGATCGGTG CGCGGGAGGG  
GAGCTAGAAA CCGGAACCAT CAAACCCACC CGCTCTCGCC GAAGCAGCGG GTCTAGCCAC GCCTCCCTCCC

BamHI  
~~~~~

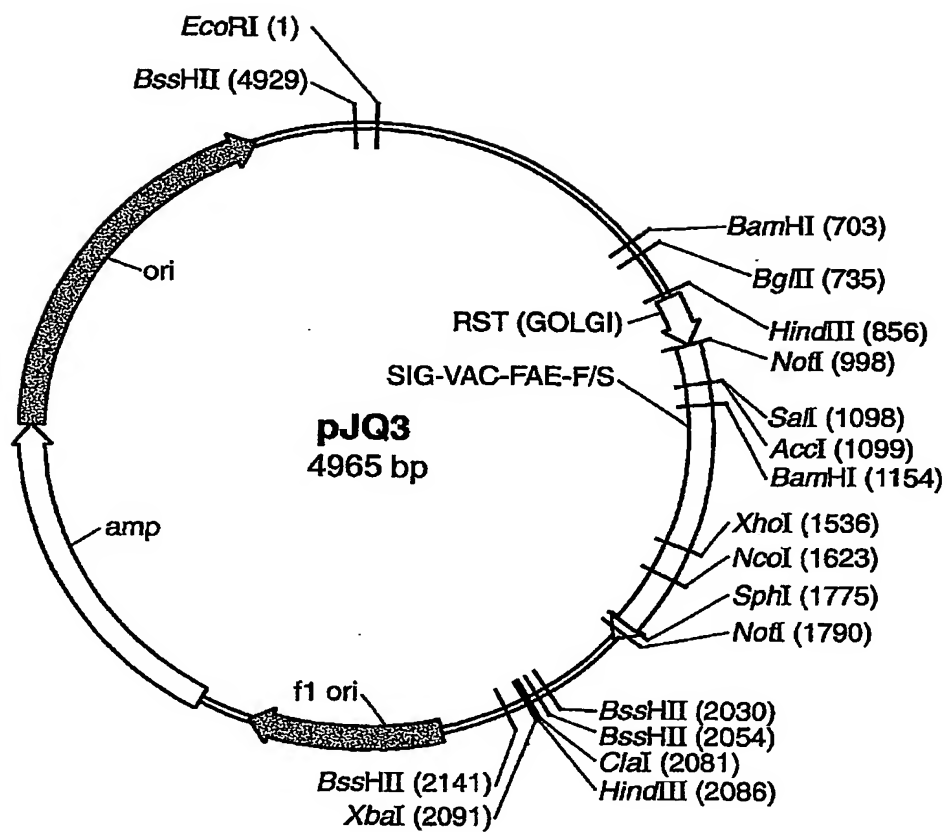
4971 GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTCGGCC CGGATCCTCG CCGGGAATGG GGCTCTCGGA
CGCCCTAGAG CGCCGACCGC AGAGGCCCGC ACTCAGCCCG GCCTAGGAGC GCCCTTACC CCGAGAGCCT

BglII
~~~~~

5041 TGATAGATCTT CTTTCTTTCT TCTTTTGTG GTAGAAATTG AATCCCTCAG CATTGTTCAT CGGTAGTTTT  
ACATCTAGAA GAAAGAAAG AAAAAACAC CATCTTAAC TTAGGGAGTC GTAACAAGTA GCCATCAAAA

5111 TCTTTTCATG ATTGTGACA AATGCAGCCT CGTCCGGAGC TTTTGTGTAG GTAG  
AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCCTCG AAAAAACATC CATC

FIG. 45H

**FIG. 46A**

**EcoRI**  
 1 AATCCACAA TGAACAATAA TAAGATTAAA ATAGCTTGCC CCGTTGACAG CGATGGGTAT TTTTCTAGT  
 TTAAGGTGTT ACTTGTTATT ATCTAATTT TATCGAACGG GGGCAACGTC GCTACCCATA AAAAGATCA  
 71 AAAATAAAAG ATAACTTAG ACTCAAAACA TTTACAAAAA CAACCCCTAA AGTCCATAAG CCCAAAGTGC  
 TTTTATTTTC TATTGAATC TGAGTTTGT AAATGTTTTT GTTGGGATTC TCAGGATTTT GGGTTTCACG  
 141 TATGCACGAT CCATAGCAAG CCCAGCCCAA CCCAACCCAC CCCAGTGCAG CCAACTGGCA  
 ATACGTGCTA GGTATCGTTC GGGTCGGGTT GGGTTGGGTT GGGTCACGTC GGTTCACCGT  
 211 AATAGTCTCC ACCCCGGCA CTATCACCGT GAGTTGTCCG CACCAACGCA CGTCTCGCAG CCAAAAAA  
 TTATCAGAGG TGGGGCCCGT GATAGTGGCA CTCAACAGGC GTGGTGGCGT GCAGAGCGTC GGTTTTTTTT  
 281 AAAAAGAAAG AAAAAAAGA AAAAAGAAA CAGCAGGTGG GTCCGGGTCTG TGGGGGCCGG AAAAGCGAGG  
 TTTTCTTTC TTTTCTTCT TTTTCTTCTT GTCTCTCCAC CAGGCCCAGC ACCCCCGGCC TTTTCGCTCC  
 351 AGGATCGCGA GCAGCGACGA GGCCCGGCC TCCCTCCGCT TCCAAAGAAA CGCCCCCAT CGCCACTATA  
 TCCTAGCGCT CGTCGTGCT CCGGGCCGG AGGAGAGCGA AGGTTCTTTT GCGGGGGTA GCGTGTATAT  
 421 TACATACCC CCCCTCTCT CCCATCCCC CAACCTTACC ACCACACCA CCACCACTC CTCCTCCCTC  
 ATGTATGGG GGGAGAGGA GGTAGGGGG GTTGGGATGG TGGTGTGGT GGTGGTGGAG GAGGGGGGAG  
 491 GGTGCCGGAC GACGAGCTCC TCCCCCTCC CCTCCGCGG CCGCCGGTAA CCACCCCGCC CCTCTCTCT  
 CGACGGCTG CTGCTCGAG AGGGGGGAG GGGAGGCGGC GCGGCCCATT GGTGGGGCGG GAGAGGAGA  
 561 TTTCTTCTCC GTTTTTTTT TCGTCTCGGT CTCGATCTTT GGCCTTGGTA GTTTGGGTGG GCGAGAGCGG  
 AAGAAAGAG CAAAAAAA AGCAGAGCCA GAGTAGAAA CCGGAACCAT CANACCCACC CGCTCTCACC  
 631 CTTCTCTGCC CAGATCGGTG GCGGGAGGG GCGGGATCTC GCGGCTGGCG TCTCCGGGCG TGAGTGGCG  
 GAAGCAGCGG GTCTAGCCAC GCGCCCTCCC GCGCCTAGAG CCGCGACCGC AGAGGCCCGC ACTCAGCCCG

**BamHI**  
 701 CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGATGATCTT CTTTCTTCT TCTTTTGTG GTAGAATTG  
 GCCTAGGAGC GCCCCTTACC CCGAGAGCCT ACATCTAGAA GAAAGAAAGA AGAAAAACAC CATCTTAAC

**BglII**  
 701 CGGATCCTCG CGGGGAATGG GGCTCTCGGA TGATGATCTT CTTTCTTCT TCTTTTGTG GTAGAATTG  
 GCCTAGGAGC GCCCCTTACC CCGAGAGCCT ACATCTAGAA GAAAGAAAGA AGAAAAACAC CATCTTAAC

**FIG. 46B**

```

771 AATCCCTCAG CATGTTTCAT CGGTAGTTTT TCTTTTCATG ATTGTGACA AATGAGCCTT CGTCCGGAGC
TTAGGGAGTC GTAACAAGTA GCCATCAAAA AGAAAAGTAC TAAACACTGT TTACGTCGGA GCACGCCTCG

HindIII
~~~~~
841 TTTTGTGTAG GTAGAAGCTT ACCATGATCC ACACCAACCT CAAAAGAGG TTCTCCCTCTT TCATCCTCGT
AAAAAACATC CATCTTCGAA TGGTACTAGG TGTGGTTGGA GTTTTCTTTC AAGAGGGAGA AGTAGGAGCA

911 CTTCTCTCTC TTCGCCGTGA TCTGGGTGTG GAAGAAGGGC TCCGACTACG AGGCCCTCAC CCTCCAAGCC
GAAGGAGGAG AAGCGGCACT AGACGCACAC CTTCTTCCCG AGGCTGATGC TCCGGGAGTG GGAGGTTCCG

NotI
~~~~~
981 AAGGAGTTC AAATGGCGGC CGCTCCACG CAGGGCATCT CCGAAGACCT CTACAGCCCTT TTAGTCGAAA
TTCTCTCAAGG TTTTACCGCG GCGGAGGTGC GTCCCGTAGA GGCTTCTGGA GATGTCGGCA AATCAGCTTT

Sall
~~~~~
1051 TGGCCACTAT CTCCCAAGCT GCCTACGCCG ACCTGTGCAA CATTCGGTCG ACTATTATCA AGGGAGAGAA
ACCGGTGATA GAGGGTTGGA CGGATGCGGC TGGACACGTT GTAAGGCAGC TGATAATAGT TCCCTCTCTT

AccI
~~~~~
BamHI
~~~~~
1121 AATTACAAT TCTCAAACTG ACATTAACGG ATGGATCCTC CGCGACGACA GCAGCAAAGA AATAATCACC
TTAAATGTTA AGAGTTTGAC TGTAATTGCC TACCTAGGAG GCGCTGCTGT CGTCGTCTCT TTATTAGTGG

1191 GTCTTCCGTG GCACTGGTAG TGATACGAAT CTACAACCTG ATACTAACTA CACCCCTCAC CCTTTTCGACA
CAGAAGGCAC CGTGACCATC ACTATGCTTA GATGTTGAGC TATGATTGAT GTGGGAGTGC GGAAAGCTGT

1261 CCTTACCACA ATGCAACGGT TGTGAAGTAC ACGGTGGATA TTATATTGGA TGGGTCTCCG TCCAGGAGCA
GGGATGGTGT TACGTTGCCA ACACCTCATG TGCCACCTAT AATATAACCT ACCCAGAGGC AGGTCTCTGT

```

FIG.\_46C

```

1331 AGTCGAGTCG CTTGTCAAAC AGCAGGTTAG CCAGTATCCG GACTACGCGC TGACCGTGAC CGGCCACGCC
 TCAGCTCAGC GAACAGTTTG TCGTCCAATC GGTCATAGGC CTGATGCGCG ACTGGCACTG GCCGGTQMGG

1401 CTCGGCGCCT CCCTGGCGGC ACTCACTGCC GCCAGCTGT CTGCGACATA CGACAACATC CGCCTGTACA
 GAGCCGCGGA GGGACCGCCG TGAGTGACGG CGGTCGACA GACGCTGTAT GCTGTTGTAG GCGGACATGT

 XhoI
                                ~~~~~
1471 CCTTCGGCGA ACCGCGCAGC GGCAATCAGG CCTTCGCGTC GTACATGAAC GATGCCTTCC AAGCCTCGAG
    GGAAGCCGCT TGGCGCGTCG CCGTTAGTCC GGAAGCGCAG CATGTACTTG CTACGGAAGG TTCGGAGCTC

1541 CCCAGATACG ACCGAGTATT TCCGGGTCAC TCATGCCAAC GACGGCATCC CAACCTGCC CCCGGTGGAG
    GGGTCTATGC TCGTCAATAA AGGCCCAGTG AGTACGGTTG CTGCCGTAGG GTTTGGACGG GGGCCACCTC

                                NcoI
                                ~~~~~
1611 CAGGGGTACG CCCATGGCGG TGTAGAGTAC TGGAGCGTTG ATCCTTACAG CGCCCAAGAAC ACATTGTCT
 GTCCCATGTC GGGTACCGCC ACAFTCATG ACCTCGCAAC TAGGAATGTC GCGGCTCTTG TGTAAACAGA

1681 GCACTGGGGA TGAAGTGCAG TCGTGTGAG CCCAGGGCGG ACAGGGTGTG AATAATGCGC ACACGACTTA
 CGTGACCCCT ACTTCACGTC ACGACACTCC GGGTCCCBC TGTCCCACAC TTATTACGGG TGTGCTGAAT

 SphI
                                ~~~~~
                                NotI
                                ~~~~~
1751 TTTTGGGATG ACGAGCGGCG CATGCACCTG GCCGCTCGCG GCCCGGAAA CCACTGAAGG ATGAGCTGTA
 AAAACCCCTAC TGCTCGCCGC GTACCTGGAC CGGCCAGCGC CGGCGCCCTTT GGTGACTTCC TACTCGACAT

1821 AAGAAGCAGA TCGTTCAAAC ATTTGGCAAT AAAGTTTCTT AAGATTGAAT CCTGTGCGCG GTCTTGCAT
 TTCTTCGTCT AGCAAGTTTG TAAACCGTTA TTTCAAAGAA TTCTAACTTA GGACAACGGC CAGAACGCTA

1891 GATTATCATA TAATTTCTGT TGAATTACGT TAAGCATGTA ATAAATTAACA TGTAAATGCAT GACGTTATTT
 CTAATAGTAT ATTAAGACA ACTTAATGCA ATTGCTACAT TATTAAATGT ACATTACGTA CTGCAATAAA

```

FIG. 46D

```

1961 ATGAGATGGG TTTTATATGAT TAGAGTCCCG CAATTATACA TTTAATACGC GATAGAAAAC AAAATATAGC
 TACTCTACCC AAAAATACTA ATCTCAGGGC GTTAATATGT AAATTATGCG CTATCTTTTG TTTTATATCG
 BssHII
      ~~~~~
      XbaI
      ~~~~~
2031 GCGCAAACTA GGATAAATTA TCGCGCGCGG TGTCACTCTAT GTTACTAGAT CGATAAGCTT CTAGAGCGGC
 CGCGTTTGAT CCTATTTAAT AGCGCGCGCC ACAGTAGATA CAATGATCTA GCTATTGAA GATCTCGCCG
 BssHII
      ~~~~~
      ClaI  HindIII
      ~~~~~
2101 CGGTGGAGCT CCAATTCGCC CTATAGTGAG TCGTATTACG CGCGCTCACT GGCGTCTGTT TTACAACGTC
 GCCACCTCGA GGTAAAGCGG GATATCACTC AGCATAATGC GCGCGAGTGA CCGGCAGCAA AATGTTGCAG
2171 GTGACTGGGA AAACCCCTGGC GTTACCCAAC TTAATCGCCT TGCAGCACAT CCCCTTTTCG CCAGCTGGCG
 CACTGACCCT TTTGGGACCG CAATGGGTTG AATTAGCGGA ACGTCGTGTA GGGGAAAAGC GGTCTGACCG
 BssHII
      ~~~~~
2241  TAATAGCGAA  GAGGCCCGCA  CCGATCGCCC  TTCCCAACAG  TTGCGCAGCC  TGAATGGCGA  ATGGGACGCG
      ATTATCGCTT  CTCCGGGGCGT  GGCTAGCGGG  AAGGGTTGTC  AACGCGTCGG  ACTTACCGCT  TACCCCTGCGC
2311  CCTGTAGCG  GCGCAATTAAG  CGCGCGGGGT  GTGGTGGTTA  CGCGCAGCGT  GACCGCTACA  CTTGCCAGCG
      GGGACATCGC  CGCGTAATTC  GCGCGGCCCA  CACCACCAAT  GCGCGTCGCA  CTGGCGATGT  GAACGGTTCG
2381  CCTAGCGCC  CGCTCCTTTC  GCTTTCCTCC  CTTCCCTTCT  CGCCACGFTC  GCCGGCTTTC  CCCGTCGAAG
      GGGATCGCGG  GCGAGGAAAG  CGAAGAAGG  GAAGGAAGA  GCGGTGCAAG  CCGCCGAAAG  GGGCAGTTCG
2451  TCTAAATCGG  GGGCTCCCTT  TAGGGTTCCG  ATTTAGTGCT  TTACGGCACC  TCGACCCCAA  AAAACTTGAT
      AGATTTAGCC  CCCGAGGGAA  ATCCCAAGGC  TAAATCAGCA  AATGCCGTGG  AGCTGGGGTT  TTTTGAACTA
2521  TAGGGTGATG  GTTCACGTAG  TGGGCCATCG  CCCTGATAGA  CGGTTTTCG  CCCTTTGACG  TTGGAGTCCA
      ATCCCACTAC  CAAGTGCATC  ACCCGGTAGC  GGGACTATCT  GCCAAAAGC  GGGAAACTGC  AACCTCAGGT

```

**FIG. 46E**

2591 CGTCTTTTAA TAGTGGACTC TTGTTCCAAA CTGGAACAAC ACTCAACCCCT ATCTCGGTCT ATTCTTTTGA  
GCAAGAAATT ATCACCCTGAG AACAAGGTTT GACCTTGTTG TGAGTTGGGA TAGAGCCAGA TAAGAAACT

2661 TTTATAAAGGG ATTTTGCCGA TTTCGGCCTA TTGGTTAAAA AATGAGCTGA TTTAAACAAA ATTAAACGCG  
AAATATTCCC TAAACGGCTT AAAGCCGGAT AACCAATTTT TTAATCGACT AAATGTGTTT TAAATTCGCG

2731 AATTTTAACA AAATATTAAAC GCTTACAATT TAGGTGGCAC TTTTTCGGGA AATGTGCGCG GAACCCCTAT  
TTAAAATTGT TTTATAATTG CGAATGTTAA ATCCACCGTG AAAAGCCCTT TTACACGCGC CTTGGGGATA

2801 TTGTTTATTT TTCTAAATAC ATTCAAAATAT GTATCCGCTC ATGAGACAAT AACCTTGATA AATGCTTCAA  
AACAAATAAA AAGATTATG TAAATTATATA CATAGGCGAG TACTCTGTTA TTGGGACTAT TTACGAAGTT

2871 TAATATTGAA AAAGGAAGAG TATGAGTATT CAACATTTCC GTGTGCGCCT TATTCCTTTT TTTGCGGCGAT  
ATTATTAATT TTTCTTTCTC ATACTCATAA GTTGTAAAGG CACAGCGGA ATAAAGGAAA AAACGCCGTA

2941 TTTGCCCTCC TGTTTTTGCT CACCCAGAAA CGCTGGTGAA AGTAAAAGAT GCTGAAGATC AGTTGGGTGC  
AAACGGAAGG ACMAAAACGA GTGGGTCTTT GCGACCACTT TCATTTTCTA CGACTTCTAG TCAACCCACG

3011 ACGAGTGGGT TACATCGAAC TGGATCTCAA CAGCGGTAAG ATCCTTGAGA GTTTTCGCCC CGAAGAACGT  
TGCTCACCCA ATGTAGCTTG ACCTAGAGTT GTCGCCATTG TAGGAACCTT CAAAAGCGGG GCTTCTTGCA

3081 TTTCCAATGA TGAGCACTTT TAAAGTTCTG CTATGTGGCG CGGTATTATC CCGTATTGAC GCCGGGCAAG  
AAAGGTTACT ACTCGTGAAA ATTTCAAGAC GATACACCGC GCCATAATAG GGCATAACTG CGGCCCGTTC

3151 AGCAACTCGG TCGCCGSCATA CACTATTCTC AGAATGACTT GGTGAGTAC TCACCAGTCA CAGAAAAAGCA  
TCGTTGAGCC AGCGGCGTAT GTGATAAGAG TCTTACTGAA CCAACTCATG AGTGGTCAGT GTCTTTTTCGT

3221 TCTTACGGAT GGCATGACAG TAAGAGAATT ATGCAGTGCT GCCATAACCA TGAGTGATAA CACTGCGGCG  
AGAATGCCTA CCGTACTGTC ATTCTCTTAA TACGTCACGA CGGTATTGGT ACTCACTATT GTGACGCGCGG

3291 AACTTACTTC TGACACGAT CGGAGGACCG AAGGAGCTAA CCCTTTTTT GCACAACATG GGGGATCATG  
TTGAATGAAG ACTGTTGCTA GCCTCCTGGC TTCTCTCGATT GGCGAAAAAA CGTGTGTAC CCCCTAGTAC

3361 TAACTCGCCT TGAATCGTTGG GAACCGGAGC TGAATGAAGC CATACCAAAC GACGAGCGTG ACACCCAGAT  
ATTGAGCGGA ACTAGCAACC CTTGGCCTCG ACTTACTTCG GTATGGTTTG CTGCTCGCAC TGTGTTGCTA

**FIG.\_46F**



3431 GCCTGTAGCA ATGGCAACAA CGTTGCGCAA ACTATTAACT GGCGAAGTAC TTACTCTAGC TTCCCGGCAA  
CGGACATCGT TACCGTTGTT GCAACGCGTT TGATAATTGA CCGCTTGATG AATGAGATCG AAGGGCCGTT

3501 CAATTAATAG ACTGGATGA GCGGATATAA GTTGCAGGAC CACTTCTGCG CTCGGCCCTT CCGGCTGGCT  
GTTAATTATC TGACCTACCT CCGCCTATTT CAACGTCTTG GTGAAGACGC GAGCCGGGAA GGCCGACCGA

3571 GGTTTATTGC TGATAAATCT GGAGCCGGTG AGCGTGGGTC TCGCGGTATC ATTGCAGCAC TGGGGGCCAGA  
CCAAATAACG ACTAATTAGA CCTGGGCCAC TCGCACCCAG AGCGCATAG TAACGTCTG ACCCCGCTCT

3641 TGGTAAGCCC TCCCGTATCG TAGTTATCTA CACGACGGGG AGTCAGGCAA CTATGGATGA ACGAAATAGA  
ACCATTCGGG AGGCAATAGC ATCAATAGAT GTGCTGCCCC TCAGTCCGTT GATACCTACT TGCTTTATCT

3711 CAGATCGCTG AGATAGGTGC CTCACTGATT AAGCATTGGT AACGTCTACA CCAAGTTTAC TCATATATAC  
GTCTAGCGAC TCTATCCACG GAGTGACTAA TTCGTAAACCA TTGACAGTCT GGTTCAAATG AGTATATATG

3781 TTTAGATTGA TTTAAACTTT CATTTTAAAT TTAAAAGGAT CTAGGTGAAG ATCCTTTTTG ATAATCTCAT  
AAATCTAACT AAATTTTGAA GTAAAAATTA AATTTTCTA GATCCACTTC TAGGAAAAAC TATTAGAGTA

3851 GACCAAAATC CCTAACGTC AGTTTTCGTT CCACTGAGCG TCAGACCCCG TAGAAAAGAT CAAAGGATCT  
CTGGTTTTAG GGAATTGCAC TCAAAAGCAA GGTGACTCGC AGTCTGGGGC ATCTTTTCTA GTTTCCTAGA

3921 TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAA ACCACCGCTA CCAGCGGTGG  
AGAACTCTAG GAAAAAAGA CCGCATTAG ACGACGAACG TTTGTTTTTT TGGTGGCGAT GGTCCGCCACC

3991 TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTTCGGAA GGTAACTGGC TTCAGCAGAG CGCAGATACC  
AAACAAACGG CCTAGTTCTC GATGGTTGAG AAAAAGGCTT CCATTGACCG AAGTCGTCTC GCGTCTATGG

4061 AAATACTGTC CTCTAGTGT AGCCGTAGTT AGGCCACCAC TTCAAGAAGT CTGTAGCACC GCCTACATAC  
TTTTATGACAG GAAGATCACA TCGGCATCAA TCCGGTGGTG AAGTCTTGA GACATCGTGG CGGATGTATG

4131 CTCGCTCTGC TAATCCTGTT ACCAGTGGCT GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT  
GAGCGAGACG ATTAGGACAA TGGTCAACCGA CGACGGTCA CCGTATTCTAG CACAGAATGG CCCAACCTGA

4201 CAAGACGATA GTTACCGGAT AAGGCGCAGC GGTCTGGGCTG AACGGGGGT TCGTGCACAC AGCCAGCTT  
GTTCTGCTAT CAATGGCCTA TTCCGCGTCTG CCAGCCCGAC TTGCCCCCA AGCACGTGTG TCGGGTCTGA

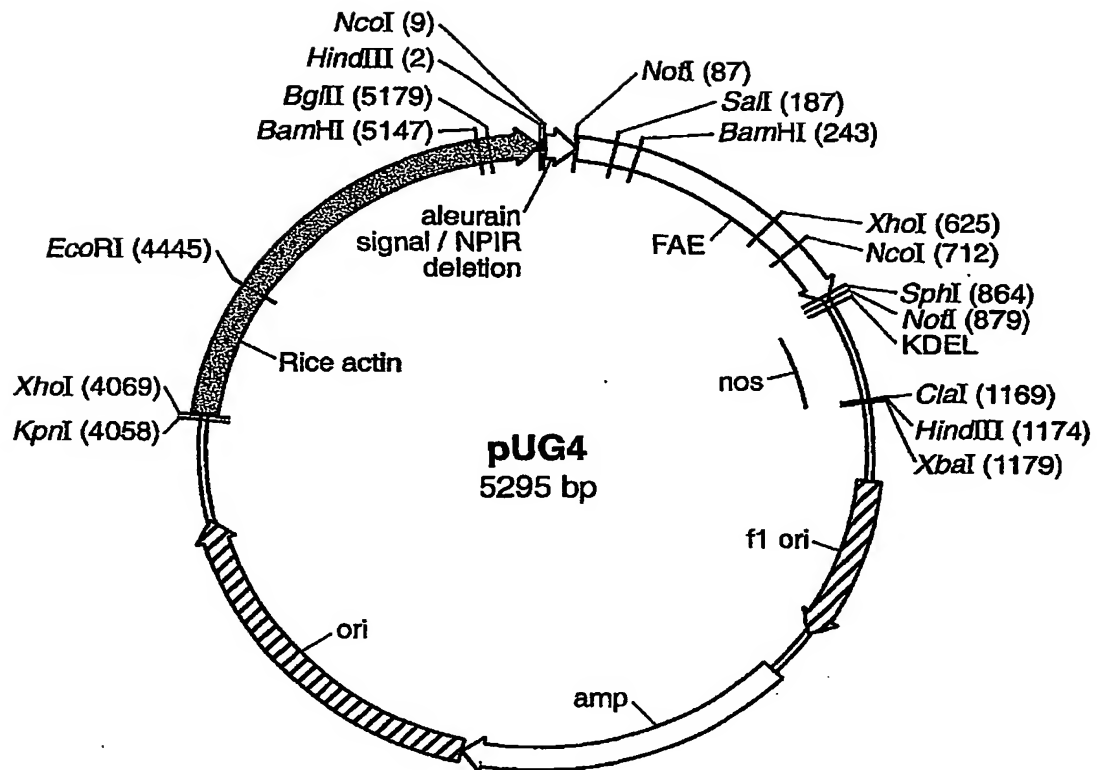
**FIG.-46G**

4271 GGAGCGAACG ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA  
 CCTCGCTTGC TGGATGTGGC TTGACTCTAT GGAATGTGCA CTCGATATC TTTCGCGGTG CGAAGGGCTT  
 4341 GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTGC GAACAGGAGA GCGCACGAGG GAGCTTCCAG  
 CCCTCTTTCC GCCTGTCCAT AGGCCATTG AGCCCAAGC CTTGTCTCTT CCGTGTCTCC CTCGAAGGTC  
 4411 GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTTCG CCACCTCTGA CTTGAGCGTC GATTTTGTG  
 CCCCTTTGCG GACCATAGAA ATATCAGGAC AGCCCAAGC GGTGGAGACT GAACTCGCAG CTAAAAACAC  
 4481 ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC AACGGGCGCT TTTTACGGTT CCTGGCCCTT  
 TACGAGCAGT CCCCCCGCCT CGGATACCTT TTTCGCGTGC TTGCGCCGGA AAATGCCAA GGACCCGGAA  
 4551 TGCTGGCCCT TTGCTCACAT GTTCTTTCTT GCGTTATCCC CTGATTTCTGT GGATAACCGT ATTACCGCCT  
 ACGACCGGAA AACGAGTGTA CAAGAAAGGA CGCAATAGGG GACTAAGACA CCTATTGGCA TAATGGCGGA  
 4621 TTGAGTGAGC TGATACCGCT CGCCGCGACC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCGGA  
 AACTCACTCG ACTATGGCGA GCGGCGTCGG CTTCCTGGCT CGCGTCGCTC AGTCACTCGC TCCTTTCGCC  
 4691 AGAGCGCCCA ATACGCAAC CGCCTCTCCC CGCGCGTTGG CCGATTCATT AATGCAGCTG GCACGACAGG  
 TCTCGCGGGT TATGCGTTTG GCGGAGAGGG GCGCGCAACC GGCTAAGTAA TTACGTCGAC CGTGTGTCTC  
 4761 TTTCCCGACT GGAAGCGGG CAGTGAGCGC AACGCAATTA ATGTGAGTTA GCTCAGTCAT TAGGCACCCC  
 AAAGGCGTGA CCTTTCGCC GTCACTCGCG TTGCGTTAAT TACACTCAAT CGAGTGAGTA ATCCGTGGGG  
 4831 AGGCTTTTACA CTTTATGCTT CCGGCTCGTA TGTGTGTGTG AATTGTGAGC GGATAACAAT TTCACACAGG  
 TCCGAAATGT GAAATACGAA GGCCGAGCAT ACAACACACC TTAACACTCG CCTATTGTTA AGTGTGTCTC

B8SHII  
 ~~~~~  
 ECoRI

4901 AAACAGCTAT GACCATGATT ACGCCAAGCG CGCAATTAAC CCTCACTAAA GGAACAAAA GCTGG
 TTTGTGCGATA CTGGTACTAA TCGGTTTCGC GCGTTAATTG GGAGTGATTT CCCTTGTTTT CGACC

FIG. 46H

**FIG. 47A**

```

      NcoI
      ~~~~~
HindIII
~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGC CCGGTCCTC CTCTGGCGC TGGCCGTGCT GGCCACGGCC GCGTCGCCG

      ~~~~~
      NotI
      ~~~~~
      . A S S R A A A S T Q G I S E D L Y S R L V E M .
71 TCGCTTCTC CCGCGGCGC GCCTCCACGC AGGGCATCTC CGAAGACCTC TACAGCCGTT TAGTCGAAT

      ~~~~~
      Sall
      ~~~~~
      . A T I S Q A A Y A D L C N I P S T I I K G E K
141 GGCCACTATC TCCCAAGCTG CCTACGCCGA CCTGTGCAAC ATTCCGTCGA CTATTATCAA GGGAGAGAAA

      ~~~~~
      BamHI
      ~~~~~
      I Y N S Q T D I N G W I L R D D S S K E I I T V
211 ATTACAATT CTCAACTGA CATTACGGA TGGATCCTCC GCGACGACAG CAGCAAAGAA ATAATCACCG
      . F R G T G S D T N L Q L D T N Y T L T P F D T .
281 TCTTCCGTGG CACTGGTAGT GATACGAATC TACAACCTGA TACTAATAC ACCCTCACGC CTTTCGACAC
      . L P Q C N G C E V H G G Y Y I G W V S V Q D Q
351 CCTACCACAA TGCACGGTT GTGAAGTACA CGGTGGATAT TATATTGGAT GGGTCTCCGT CCAGGACCAA
      V E S L V K Q Q V S Q Y P D Y A L T V T G H K L
421 GTCGAGTCGC TTGTCAACA GCAGGTTAGC CAGTATCCGG ACTACGGCT GACCGTGACC GGCCACKCCC
      . G A S L A A L T A A Q L S A T Y D N I R L Y T .
491 TCGGCGCCTC CCTGGCGCA CTCACCTGCC CCCAGCTGTC TCGGACATAC GACACATCC GCCTGTATAC
      XhoI
      ~~~~~
      . F G E P R S G N Q A F A S Y M N D A F Q A S S
561 CTTGGGCGAA CCGCGCAGCG GCAATCAGGC CTTGCGGTCG TACATGAACG ATGCTTCCA AGCCTCGAGC
      P D T T Q Y F R V T H A N D G I P N L P P V E Q
631 CCAGATACGA CGCATATTT CCGGGTCACT CATGCCAAGC ACGCATCCC AAACCTGCC CCGGTGGAGC

      ~~~~~
      NcoI
      ~~~~~

```

FIG._47B

. G Y A H G G V E Y W S V D P Y S A Q N T F V C .
 701 AGGGGTACGC CCATGGCGGT GTAGAGTACT GGAGCGTTGA TCCTTACAGC GCCCAGAACCA CATTGTCTCG
 . T G D E V Q C C E A Q G G Q G V N A H T T Y
 771 CACTGGGGAT GAAGTGCAGT GCTGTGAGGC CCAGGCGGGA CAGGGTGTGA ATAATGCGCA CACGACTTAT
 NotI
 SphI
 F G M T S G A C T W P V A A A E P L K D E L *
 841 TTGGGATGA CGAGCGGCG ATGCACCTGG CCGTCCGGG ACTGAAGGAT GAGCTGTAAA
 911 GAAGCAGATC GTTCAAAACAT TTGGCAATAA AGTTTCTTAA GATTGAATCC TGTTCGCCGT CTTGCCGATGA
 981 TTATCATATA ATTCTGTGTG AATTACGTTA AGCATGTAAT AATTAAACATG TAATGCAATGA CGTTATTAT
 1051 GAGATGGGTT TTTATGATTA GAGTCCCGCA ATTATACATT TAATACGCGA TAGAAACAA AATATAGCGC
 HindIII
 ClaI XbaI
 1121 GCAAACTAGG ATAAATATC GCGCGCGGTG TCATCTATGT TACTAGATCG ATAGCTTCT AGAGCGGCGG
 1191 GTGGAGCTCC AATTCGCCCT ATAGTGAATC GTATTACGGG CGTCACATGG CCGTCGTTT ACAACGTCGT
 1261 GACTGGGAAA ACCCTGGCGT TACCCAACTT AATCGCCTTG CAGCACATCC CCCTTTCGCC AGCTGGCGTA
 1331 ATAGCGAAGA GCGCCGACAC GATCGCCTT CCCAACAGTT GCGCAGCCTG AATGCGCAAT GGCACGCGCC
 1401 CTGTAGCGGC GCATTAAAGC GCGCGGCTG GTTGGTTACG CGCAGCGTGA CCGCTACACT TGCACGCGCC
 1471 CTAGCGCCCG CTCCTTTCGC TTTCTTCCTT TCCCTTCTCG CCACGTTCCG CCGCTTTCCT CGTCAAGCTC
 1541 TAAATCGGGG GCTCCCTTTA GGGTTCCGAT TTAGTGTCTT ACGGCACCTC GACCCCAAA AACTTGATTA
 1611 GGGTGATGGT TCACGTAGTG GGCATTCGCC CTGATAGACG GTTTTTCGCC CTTTGACGTT GAGTCCACG
 1681 TTTCTTAAAT GTGGACTCTT TCGGCTATT TCGTAAATAA TGAGCTGATT TAACAAAAAT TTAACGCGAA
 1751 TATAAGGGAT TTTGCGGATT TTACAAATTA GGTGCGACTT TTTCGGGAAA TGTGCGCGGA ACCCTATT
 1821 TTTTAACAAA ATATTACGC TTACAAATATG ATCCGCTCAT GAGACAAFAA CCTGATAAA TGCCTCAATA
 1891 GTTTATTTTT CTAATACAT TCAATATATG ACATTTCCGT GTCCGCCCTTA TTCCCTTTT TCGGGCATTT
 1961 ATATTGAAA AGGAAGAGTA TGAGTATCA CCCAGAAACG CTGGTGAAG CTTGAGAGT TTTCCGCCCG AAGAACGTTT
 2031 TGCCTTCCTG TTTTTCCTCA CATCGAACTG GATCTCAACA CCGGTAAGAT CTTGAGAGT TTTCCGCCCG AAGAACGTTT
 2101 GAGTGGGTTA AGCACTTTTA AAGTCTGCT ATGTGGCGCG GTATTATCCC GTATTGACGC CGGGCAAGAG
 2171 TCCAATGATG GCGCATACA CTATTCTCAT GATGAGTGG CATTACCATG AGTGAATAACA CTGCGGCCAA
 2241 CAACCTCGTC GCGCATACA CTATTCTCAT GATGAGTGG CATTACCATG AGTGAATAACA CTGCGGCCAA
 2311 TTACGGATGG CATGACAGTA AGAGATTAT GAGGCTAAC CCATTTTTCG ACAACATGGG GATCATGTA
 2381 CTACTTCTG ACAACGATCG GAGGACCGAA GAGGCTAAC CCATTTTTCG ACAACATGGG GATCATGTA
 2451 ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA TACCAACGGA CGAGCGTGAC ACCACGATGC
 2521 CTGTAGCAAT GGCAACAACG TTGCGCAAAAC TATTAACTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA

FIG.-47C

2591 ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA CTTCCTGCGCT CGCCCTTCC GGCTGGCTGG
 2661 TTTATTGCTG ATAAATCTGG AGCGGTGAG CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG
 2731 GTAAGCCCTC CCGTATCGTA GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGARC GAAATAGACA
 2801 GATCGCTGAG ATAGGTGCCCT CACTGANTAA GCATTGGTAA CTGTGAGACC AGTTTACTC ATATATACCT
 2871 TAGATTGATT TAAAACCTCA TTTTAAATTT AAAAGGATCT AGGTGAAGAT CCTTTTGTG ATCTCATGA
 2941 CCAAATCCC TTAACGTGAG TTTTCTCTGC GCGTAATCTG ACCAATCTTT TTTCCGAAAG TAACGTGCTT CAGATACCAA
 3011 TGTGTGCCG ATCAAGAGCT TTTTCTCTGC GCGTAATCTG ACCAATCTTT TTTCCGAAAG TAACGTGCTT CAGATACCAA
 3081 ATACTGTCTT TCTAGTGTAG CCGTAGTTAG GCGTAGTTAG GCGTAGTTAG GCGTAGTTAG GCGTAGTTAG GCGTAGTTAG
 3151 CGCTCTGCTA ATCCTGTTAC CAGTGGCTGC TCGCAGTGGC GCGTAGTTAG GCGTAGTTAG GCGTAGTTAG GCGTAGTTAG
 3221 AGACGATAGT TACCGGATTA GCGCGAGCGG TCGGCTGGA TCGGCTGGA TCGGCTGGA TCGGCTGGA TCGGCTGGA
 3291 AGCGAACGAC CTACACCCGAA CTGAGATACC TACAGCGTGA TACAGCGTGA TACAGCGTGA TACAGCGTGA TACAGCGTGA
 3361 GAGAAAGCG GACAGGTATC CCGTAAGCGG CAGGTCGGA CAGGTCGGA CAGGTCGGA CAGGTCGGA CAGGTCGGA
 3431 GGAACGCCCT GGTATCTTTA TAGTCTGTC GGTCTGTC GGTCTGTC GGTCTGTC GGTCTGTC GGTCTGTC
 3501 GCTCGTCAGG GGGCGGAGC CTATGGAATA TCTTCTCTGC TCTTCTCTGC TCTTCTCTGC TCTTCTCTGC TCTTCTCTGC
 3571 GCTCGTCAGG GGGCGGAGC CTATGGAATA TCTTCTCTGC TCTTCTCTGC TCTTCTCTGC TCTTCTCTGC TCTTCTCTGC
 3641 GAGTGAGCTG ATACCGCTCG CCGCAGCCGA CCGCAGCCGA CCGCAGCCGA CCGCAGCCGA CCGCAGCCGA CCGCAGCCGA
 3711 AGCGCCCAAT ACGCAACCG CCTCTCCCG CCGCAGCCGA CCGCAGCCGA CCGCAGCCGA CCGCAGCCGA CCGCAGCCGA
 3781 TCCCGACTGG AAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGCTAGC TCACTCATTA GGCACCCCGAG
 3851 GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGA TTGTGTGGA TTGTGTGGA TTGTGTGGA TTGTGTGGA
 3921
 3991 ACAGCTATGA CCAATGATTAC GCCAAGCGCG CAATTAACCC TCACATAAAGG GAACAAAAGC TGGGTACCCG
 4061 GCGCCCTCTC GAGGTCAATC ATATGCTTGA GAAGAGAGTC GGGATAGTCC AAAATAAAGC AAAGGTAAAG
 4131 TTACCTGGTC AAAAGTGAAT ACATCAGTTA AAAGGTGGA TAAGTAAAT TAAGTAAAT TAAGTAAAT TAAGTAAAT
 4201 CAAAGTGAAA TTTACTCTTT TCTACTATTA TAAAATTTGA GGTATTTTG TCGGTACTTT GATACGTCAT
 4271 TTTTGTATGA ATTGGTTTTT AAGTTTATTC GCGATTGGA AATGCATATC TGTATTGAG TCGGTTTTFA
 4341 AGTTCGTTGC TTTTGTAAAT ACAGAGGGAT TTGTATAAGA AATATCTTTA AAAAACCCAT ATGCTAATTT
 4411 GACATAATTT TTGAGAAAAA TATATATTCA GCGGAATTC ACAATGAACA ATAATAAGAT TAAATAGCT
 4481 TGCCCCCGTT GCAGCGATGG GTATTTTTTC TAGTAAATTA AAGATAAAC TTAGACTCAA AACATTACA
 4551 AAAACAACCC CTAAAGTCTT AAAGCCCAAA GTGCTATGCA CGATCCATAG CAAGCCGAGC CCAACCCAC
 4621 CCAACCCAC CCACCCAGT GCAGCCCACT GGCAATAGT CTCCACCCCG GGCACCTATCA CCCTGAGTTG

RpnI
 EcoRI

XhoI

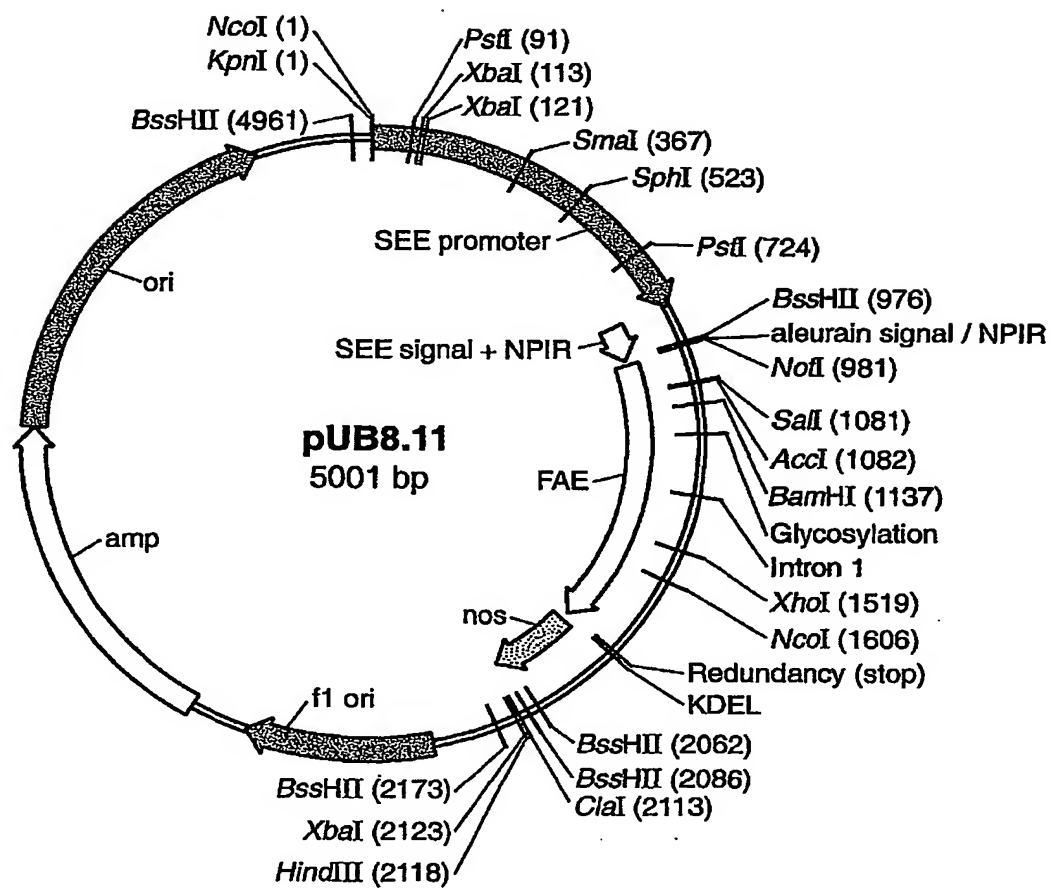
FIG. 47D

```

4691 TCCGCACCAC CGCACGTC TC GCGCCAAAAA AAGAAAAAAGA AAAACAGCAG
4761 GTGGGTCCGG GTCGTGGGGG CCGGAAAAGC GAGGAGGATC GCGAGCAGCG ACGAGGCCCG GCCCTCCCTC
4831 CGCTTCCAAA GAAACGCCCC CCATCGCCAC TATATACATA CCCCCCCTC TCCTCCCATC CCCCCAACCC
4901 TACCACCACC ACCACCACCA CCTCCTCCCC CCTCGCTGCC GGACGACGAG CTCCTCCCCC CTCCCCCTCC
4971 GCCGCCGCCG GTAACCAACC CGCCCTCTC CGCCCTTTT CTCCGTTTTT TTTTTCGTCT CGGTCTCGAT
5041 CTTTGGCCCT GGTAGTTTGG GTGGGCGAGA GCGGCTTCGT CGCCCAAGATC GTGCGCGGG AGGGGCGGGA
                                     BamHI
                                     ~~~~~
5111 TCTCGCGGCT GGCGTCTCCG GGCGTGAATC GGCCCGGATC CTCGCGGGGA ATGGGGCTCT CGGATGTAGA
                                     BglII
                                     ~~~
5181 TCTTCTTCTT TTTCTTTTTC TGTGGTAGAA TTTGAATCCC TCAGCATGCT TCATCGGTAG TTTTCTTTTT
5251 CATGATTTGT GACAAATGCA GCTTCGTGCG GAGCTTTTTT GTAGC

```

FIG.-47E

**FIG. 48A**

NcoI
 ~~~~~  
 KpnI  
 ~

1 CATGGGCCAG GTATATATAT GGGATATCTC AAGCAATAA TCGAATATC ACCATTGGCT ACAATATCTG  
 GTACCCGGTC CATATTAATA CCTATATAG CCCTATATAG TTCGTTTATT AGCTTATAG TGGTAACCGA TGTATATAGC

PstI  
 ~~~~~

71 AGTCCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGTG ATCTAGAAT CTAGATAGCA CAGCCACAGC
 TCGAGGCTCA AGACTGACGT CAGACCTACT GCGCACAA CA TAGATCTTGA GATCTATCGT GTCGGTGTCTG

XbaI XbaI
 ~~~~~ ~~~~~

141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG TCATCACAC CTCTGCCCTG AGAANGGATG CTCCTGACGT  
 TGGATGTCTT CACGCTGGA ACACCTGACA TCGTCAACCA TCACGCGCTC CCAACAAAAT ATCGTCCCCC

211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC  
 ACGCGGCAA CAGGTAAGGT TGCCGTAGTG AGAGTTGGTT AGTCCGCGAG GGTGTGTTTA TAGCAGGGGG

281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTGCT GAATCTCGCT TCCACTGGCC  
 TACAGAACCG CCTCTCTCTC ATGTATGTAC GACAGCGCGG CAAAACAGA CTTAGAGCGA AGGTGACCGG

SmaI  
 ~~~~~

351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCTG TCACCCCTGG CGTCATGGGA
 TTAGTCGAGT CGAGGGCCCT CGAGTGAGTA AGTTCFAGG TAGCAGCAGC AGTGGGGACC GCAGTACCTT

421 TGGAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCCAAT
 ACCFTTCTTT GGAGGCACG AGCCTACTCA GTCGGTATAG GGGCTTGTCT CATGACGTTT TATTGGGTTA

SphI
 ~~~~~

491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTGG CTTAATTGAC TTTATTTTGG  
 AGTCTAAGGG GGTATCTCTT TTCAATATCGT ACGAAAGCCC AAAACAAACC GAATTAAC TG AATATAAAC

561 TTGGAGTTGA ATGCTGATTT GTTGTGTA AA ATGCCAAC ATCTGAATAT CGAGACGGAT AATAGGCTGG  
 AACCTCAACT TACGACTAAA CAACACATTT TACGGGTGG TAGACTTATA GCTCTGCCTA TTATCCGACC

FIG. 48B

```

631  CTAATAAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
    GATTAATTAA ATATCGTTCT AAGACATCAC GTGTAGCGTT TATAGAAGA CCCGTAATGT CGACCTCCGA

    PstI
    ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
 AGTAGTCGGA CTTTGTGAGA CGTCTCGGAC TTGTTTCAAC ACTTCGCACC GCTACTCTAC CCATATTTTG

771 CCCCAGCACC GGGACGCGAG CTCCTGCTTA CCAGTACCAT CTCGCCCTGC TCCCCCTGCC GGACGACCCA
 GGGCCCGTGG CCTGCGCTC GAGGCGGAT GGTATGGTA GAGCGGAGCG AGGGGACGG CCTGCTGGGT

841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
 CATTTTATGA CAACGGGTGA CGGCCCGCTC TACCGGGTGC CGCGTAGGA GAAGAACCSC GAGCGGCAGA

 BssHII
    ~~~~~
    NotI
    ~~~

911 TGGCCACCGC CGCGGTGGCC GCCGCATCNT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
 ACCGGTGGCG GCGCCACCGG CGCGGTAGNA ACCGCTGAG GTTGGGCTAG GCCGGCAGT GGCCTGCGCG

 NotI
    ~~~~~

981  GGCCGCCCTCC ACGCAGGGCA TCTCCGAAGA CCTCTACAGC CGTTAGTGG AATGGCCAC TATCTCCCAA
    CCGCGGAGG TCGTCCCGT AGAGGCTTCT GGAGATGTCG GCAATCAGC TTATCCGGTG ATAGAGGGTT

    Sali
    ~~~~~
 AccI
    ~~~~~

1051 GCTGCCCTACG CCGACCTGTG CAACATTCCG TCGACTATTA TCAAGGGAGA GAAATTTTAC AATTCCTCAA
    CGACGGATGC GGCTGGACAC GTTGTAAAGC AGCTGATAAT AGTTCCCTCT CTTTAAATG TTAAGAGTTT

```

**FIG. 48C**

BamHI  
 ~~~~~  
 1121 CTGACATTAA CGGATGGATC CTCGGGACG ACAGCAGCAA AGAATAATC ACCGCTTCC GTGGCACTGG
 GACTGTAAAT GCCTACCTAG GAGCGCTGC TGTCTCGTT TCTTTATTAG TGGCAGAAG CACCGTGACC

 1191 TAGTGATACG AATCTACAAC TCGATACTAA CTACACCTTC ACGCCTTTC ACACCTTACC ACAATGCAAC
 ATCACTATGC TTAGATGTTG AGCTATGATT GATGTGGAG TCGGGAAGC TGTGGATGG TGTACGTTG

 1261 GGTGTGAAG TACACGGTGG ATATTATATT GGATGGTCT CCGTCCAGGA CCAAGTCGAG TCGCTTGTCA
 CCAACACTTC ATGTGCCACC TATAATATAA CCTACCCAGA GGCAGGTCCT GGTTCAGTC AGCGAACAGT

 1331 AACAGCAGGT TAGCCAGTAT CCGGACTACG CGCTGACCGT GACCGGCCAC KCCCTCGGG CCTCCCTGGC
 TTGTCTGTTCA ATCGGTCTATA GGCTGTATG CCGACTGGCA CTGGCCGGTG MGGGAGCCGC GGAGGGACCG

 1401 GGCACCTACT GCCGCCCAGC TGTCTGGAC ATACGACAAC ATCCGCTGT ACACCTTCGG CGAACCAGCGC
 CCGTGAATGA CCGCGGGTGC ACAGACGCTG TATGCTGTTG TAGCGGACA TGTGAAGCC GCTTGGCGCG

 XbaI
 ~~~~~  
 1471 AGCGGCAATC AGGCTTTCG CTCGTACATG AACGATGCTT TCCAAGCTC GAGCCAGAT ACGACGCAGT  
 TCGCCGTTAG TCCGGAAGCG CAGCATGTAC TTGCTACGGA AGGTTCCGAG CTCGGGTCTA TGCTGCGTCA  
  
 NcoI  
 ~~~~~  
 1541 ATTTCCGGGT CACTCATGCC AACGACGCA TCCCAAACCT GCCCCCGGTG GAGCAGGGGT ACGCCCATGG
 TAAAGGCCCA GTGAGTACGG TTGCTGCCGT AGGTTTGA CCGGGGCCAC CTCGTCCCA TCGGGGTACC

 1611 CGGTGTAGAG TACTGGAGCG TTGATCTTA CAGCGCCAG AACACATTG TCTGACTGG GGATGAAGTG
 GCCACATCTC ATGACCTCGC AACTAGGAA GTCCGGGGTC TTGTGTAAC AGACGTGACC CCTACTTCAC

 1681 CAGTGTGTG AGGCCCAGGG CGGACAGGCT GTGAATAATG CGCACACGAC TTATTTTGGG ATGACGAGCG
 GTCACGACAC TCCGGGTCCC GCCTGTCCA CACTTATTAC GCCTGTGCTG AATAAACCC TACTGCTCGC

 1751 GAGCCTGTAC ATGGTGATCA GTCAATTCAG CCTCCCCGAG TGTACCAGGA AAGATGGATG TCCTGGAGAG
 CTCGGACATG TACCACATAG CAGTAAAGTC GGAGGGCTC ACATGGTCTT TTCTACCTAC AGGACCTCTC

FIG. 48D

1821 GGGGCCGCGT AACCACTGAA GGATGAGCTG TAAAGAAGCA GATCGTTCAA ACATTGGCA ATAAAGTTTC
CCCCGGCGCA TTGGTGACTT CCTACTCGAC ATTCTTTCGT CTAGCAAGTT TGTAAACCGT TATTTCAAAG

1891 TTAAGATTGA ATCCTGTTGC CGGTCTTGGC GCCAGAACGC TACTAATAGT ATATTAAAGA CAACTTAATG GTTAAGCATG
AATCTAACT TAGGACAACG GCCAGAACGC TACTAATAGT ATATTAAAGA CAACTTAATG CAATTCGTAC

1961 TAATAATTAA CATGTAATGC ATGACGTTAT TTATGAGATG GGTTTTATG ATTAGAGTCC CGCAATTATA
ATTATTAAAT GTACATTACG TACTGCAATA AATACTCTAC CCAAAATAC TAATCTCAGG GCGTTAATAT

2031 CATTTAATAC GCGATAGAAA ACAAAATATA GCGCGCAAC TAGGATAAAT TATCGCGCGC GGTGTCACT
GTAAATTATG CGCTATCTTT TGTTTTATAT CGCGCGTTTG ATCCTATTTA ATAGCGCGCG CCACAGTAGA

2101 ATGTTACTAG ATCGATAAGC TTCTAGAGCG GCCGGTGGAG CTCCAATTCC CCTATAGTG AGTCGTATTA
TACAATGATC TAGCTATTCC AAGATCTCGC CGGCCACCTC GAGGTTAAGC GGGATATCAC TCAGCATAAAT

2171 CGCGCGCTCA CTGGCCGTCG TTTTACAACG TCGTGACTGG GAAAACCCCTG GCGTTACCCA ACTTAATCGC
GCGCGCGAGT GACCGGCAGC AAAATGTTGC AGCACTGACC CTTTGTGGGAC CGCAATGGGT TGAATTAGCG

2241 CTTGCAGCAC ATCCCCCTTT CGCCAGTGG CGTAATAGCG AAGAGGCCCG CACCGATCGC CCTTCCCAAC
GAACGTCGTG TAGGGGAAA GCGGTCGACC GCATTATCGC TTCTCCGGGC GTGGCTAGCG GGAAGGGTTG

2311 AGTTGCGCAG CCTGAATGGC GAATGGGAGG CGCCCTGTAG CGGCGCATTA AGCGCGCGCG GTGTGGTGGT
TCAACGCGTC GGACTTACCG CTTACCCCTGC GCGGACATC GCCGCGTAAT TCGCGCGGCC CACACCAACA

2381 TACGCGCAGC GTGACCGCTA CACTTGCCAG CGCCCTAGCG CCCGCTCCTT TCGCTTTCTT CCCTTCCCTT
ATGCGCGTCG CACTGGCGAT GTGAACGCTC GCGGATCGC GGGCGAGGAA AGCGAAGAA CCCTTCCCTT

1..
xbaI

~~~~~  
Clal HindIII

FIG. 48E

2451 CTCGCCACGT TCGCCGGCTT TCCCCGTCAA GCTCTAATC GGGGGCTCCC TTTAGGGTTC CGATTTAGTG  
 GAGCGGTGCA AGCGGCCGAA AGGGCGAGTT CGAGATTAG CCCCCGAGG AATCCCAAG GCTAAATCAC  
 2521 CTTTACGGCA CCTCGACCCC AAAAAACTTG ATTAGGGTGA TGGTTCACGT AGTGGGCCAT CGCCCTGATA  
 GAAATGCCGT GGAGCTGGGG TTTTITGAACT TAATCCCACT ACCAAGTGCA TCACCCGGTA GCGGGACTAT  
 2591 GACGGTTTTT CGCCCTTTGA CGTTGGAGTC CAGCTTCTTT AATAGTGGAC TCTTGTTCCTA AACTGGAACA  
 CTGCCAAAAA GCGGGAACCT GCAACCTCAG GTGCAAGAAA TTATCACCTG AGAACAAAGT TTGACCTTGT  
 2661 ACACTCACCC CTATCTCGGT CTATCTCTTT GATTATAGA GGATTTGGCC GATTCGGCC TATTGGTTAA  
 TGTGAGTTGG GATAGAGCCA GATAAGAAA CTAAATATTC CCTAAACCGG CTAAAGCCGG ATAAACAATT  
 2731 AAAATGAGCT GATTAAACA AAATTTAACG CGAATTTTAA CAAATATTA ACGCTTACAA TTTAGGTGGC  
 TTTTACTCGA CTAAATTGTT TTTAAATGTC GCTTAAAAAT GTTTTATAAT TCGGAATGTT AAATCCACCG  
 2801 ACTTTTCGGG GAAATGTGG CGGAACCCCT ATTTGTTTAT TTTTCTAAT ACATTCAAAT ATGTATCCGC  
 TGAAAAAGCCC CTTTACACGC GCCTTGGGGA TAAACAATA AAAAGATTIA TGTAAGTTTA TACATAGGCG  
 2871 TCATGAGACA ATAAACCTGA TAATGTCTC AATAATATG AAAAAGGAAG AGTATGAGTA TTCAACATTT  
 AGTACTCTGT TATTGGGACT ATTTACGAAG TTATATAAC TTTTCTCTTC TCATACTCAT AAGTTGTAAA  
 2941 CCGTGTGCCC CTTATTCCCT TTTTTCGGC ATTTTGCCCT CTTGTTTTTG CTCACCCAGA AACGCTGGTG  
 GGCACAGCGG GAATAAGGGA AAAACGCGG TAAAACGGA GAGTGGGTCT TTGGGACCAC  
 3011 AAAGTAAAA AGCTGAAGA TCAGTTGGGT GCACGAGTGG GTTACATCGA ACTGGATCTC AACAGCGGTA  
 TTTCAATTTT TACGACTTCT AGTCAACCCA CGTGTCTCACC CAATGTAGCT TGACCTAGAG TTGTGCCCCAT  
 3081 AGATCCTTGA GAGTTTTCGC CCCGAAGAAC GTTTTCCAAT GATGAGCAGT TTTAAAGTTC TGCTATGTGG  
 TCTAGGAACCT CTCAAAAGCG GGGCTTCTTG CAAAAGTTA CTACTCGTGA AATTTCAAG AGGATACACC  
 3151 CGCGGTATTA TCCCGTATTG ACGCCGGGCA AGAGCAACTC GGTCCCGGCA TACACTATTC TCAGAATGAC  
 GCGCCATAAT AGGGCATAAC TGCGGCCCGT TCTCGTTGAG CCAGCGGCGT ATGTGATAAG AGTCTTACTG  
 3221 TTGGTTGAGT ACTCACCAGT CACAGAAAA CATCTTACGG ATGGCATGAC AGTAAGAGAA TTATGCAGTG  
 AACCAACTCA TGAGTGGTCA GTGTCTTTTC GTAGAATGCC TACCGTACTG TCATTTCTCTT AATACGTCAC

**FIG.\_48F**

3291 CTGCCATTAAC CATGAGTGAAT AACACTGCGG CCAACTTACT TCTGACAACG ATCGGAGGAC CGAAGGAGCT  
GACGGTATTG GTACTCACTA TTGTGACGCC GTTGAATGA AGACTGTTCG TAGCCTCCTG GCTTCCTCGA

3361 AACCGCTTTT TTGCACAACA TGGGGGATCA TGTAACCTCG CTTGATCGTT GGGAAACCGGA GCTGAATGAA  
TTGGCGAAAA AACGTGTGT ACCCCCTAGT ACATTGAGCG GAACTAGCAA CCTTTGGCCT CGACTTACTT

3431 GCCATACCAA ACGACGAGCG TGACACCACG ATGCTCTGAG CAATGGCAAC AACGTTGCGC AAACATATTAA  
CGGTATGCTT TGCTGCTCGC ACTGTGTGTC TACGGACATC GTTACCGTTG TTGCAACGCG TTGTGATAATT

3501 CTGGCGAACT ACTTACTCTA GCTTCCCAGC AACAAATTAAT AGACTGAGT GAGGCGGATA AAGTTGCAAG  
GACCGCTTGA TGAATGAGAT CGAAGGGCCG TTGTTAATTA TCTGACCTAC CTCGCGCCTAT TTCAACGTC

3571 ACCACTTCTG CGCTCGGCC TCCCGGCTGG CTGGTTTATT GCTGATAAAT CTGGAGCCGG TGAGCGTGGG  
TGGTGAAGAC GCGAGCCGGG AAGGCCGACC GACCAATAAA CGACTATTTA GACCTCGGCC ACTCGCACCC

3641 TCTCGCGGTA TCATTGCAGC ACTGGGGCCA GATGTAAGC CCTCCCGTAT CGTAGTTATC TACACGACGG  
AGAGCGCCAT AGTAACGTG TGACCCCGGT CTACCATTCG GGAGGGCATA GCATCAATAG ATGTGCTGCC

3711 GGAGTCAGGC AACTATGGAT GAACGAAATA GACAGATCGC TGAGATAGGT GCCTCAGTGA TTAAGCATTG  
CCTCAGTCCG TTGATACCTA CTTGCTTTAT CTGTCTAGCG ACTCTATCCA CGGAGTGACT AATTCGTAAAC

3781 GTAACTGTCA GACCAAGTTT ACTCATATAT ACTTTAGATT GATTTAAAC TTTCATTTTAA ATTTAAAAAGG  
CATTGACAGT CTGGTTCAAA TGAGTATATA TGAATCTAA CTAAATTTTG AAGTAAAAAT TAAATTTTCC

3851 ATCTAGGTGA AGATCCCTTT TGATAATCTC ATGACCAAAA TCCCTTAACG TGAGTTTTCG TTCCACTGAG  
TAGATCCACT TCTAGGAAA ACTATTAGAG TACTGTTTTT AGGGAATGC ACTCAAAAGC AAGGTGACTC

3921 CGTCAGACCC CGTAGAAAAG ATCAAAGGAT CTTCTTGAGA TCCTTTTTTT CTGCGCGTAA TCTGCTGCTT  
GCAGTCTGGG GCATCTTTTC TAGTTTCCCTA GAAGAACTCT AGGAAAAAAA GACCGGCATT AGACGACGAA

3991 GCAAACAAAA AAACACCCGC TACCAGCGGT GGTTCGTTG CCGGATCAAG AGCTACCAAC TCTTTTTCGG  
CGTTGTGTTT TTGGTGCGC ATGCTCGCCA CCACAACAAC GGCCTAGTTC TCGATGGTTG AGAAAAAGGC

4061 AAGGTAACTG GCTTCAGCAG AGCGCAGATA CCAAACTAG TCCTTCTAGT GTAGCCGTAG TTAGGCCACC  
TTCCATTGAC CGAAGTCGTC TCGCGTCTAT GGTATTGAC AGGAAGATCA CATCGGCATC AATCCGGTGG

**FIG.\_48G**

4131 ACTTCAAGAA CTCGTGTAGCA CCGCCTACAT ACCTCGCTCT GCTAATCCTG TTACCAGTGG CTGCTGCCAG  
TGAAGTTCTT GAGACATCGT GCGGATGTA TGGAGGAGA CGATTAGGAC AATGGTCACC GACGACGGTC

4201 TGGCGATAAG TCGTGTCTTA CCGGGTTGGA CTCAGACGA TAGTTACCGG ATAAGGCCCA GCGGTCTGGC  
ACCGCTATTC AGCACAGAAT GSCCACAACCT GAGTCTGCT ATCAATGGCC TATTCGCGCT CGCCAGCCCC

4271 TGAACGGGGG GTTCGTGCAC ACAGCCACG TGTGCGGTG AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGATGTG  
ACTTGCCCCC CAAGCACGTG TGTGCGGTG AACCTCGCTT GCTGGATGTG GCTTGACTCT ATGGATGTG

4341 GTGAGCTATG AGAAGCGCC ACGCTTCCG AAGGAGAAA GCGGACAGG TATCCGGTAA GCGGCAGGGT  
CACTCGATAC TCTTTCGCG TCGGAAGGCT TTCCCTCTTT CCGCTGTCC ATAGGCCATT GCGCTGCCA

4411 CGGAACAGGA GAGCGCACGA GGGAGCTTCC AGGGGAAAC GCCTGGTATC TTTATAGTCC TGTCGGGTTT  
GCCTTGTCTT CTCGCGTGT CCTCGAAGG TCCCCCTTG CCGACCATAG AATATACAG ACAGCCCAA

4481 CGCCACCTCT GACTTGAGG TCGATTTTGG AGCTAAAAAC ACTACGAGCA GTCCCCCGC CTCGGATACC TTTTTCGGT  
GCGGTGGAGA CTGAACCTGC AGCTAAAAAC ACTACGAGCA GTCCCCCGC CTCGGATACC TTTTTCGGT

4551 GCAACGCGGC CTTTTTACGG TTCCCTGGCT TTTGCTGCC TTTTGTCTAC ATGTTCTTTC CTGCGTTATC  
CGTTGCGCCG GAAAAATGCC AAGGACCGGA AAACGACCG AAAACGAGTG TACAAGAAAG GACGCAATAG

4621 CCTTGATCT GTGGATAACC GTATTACCG CTTTGTAGTGA GCTGATACCG CTCGCCGCG CCGAACGACC  
GGGACTAAGA CACCTATTGG CATAATGGG GAAACTCACT CGACTATGGC GAGCGGCGTC GGCCTGCTGG

4691 GAGCGCAGCG AGTCAGTGAG CGAGGAAGCG GAAGAGCGCC CAATACGCA ACCGCTCTC CCCGCGCGTT  
CTCGCGTGGC TCAGTCACTC GCTCCTTCGC CTTCTCGCG GTTATGCGTT TGGCGGAGG GGGCGCGCAA

4761 GGCCGATTCA TTAATGCAGC TGGCAGCA GGTTCCTCGA CTGGAAGCG GGCAGTGGC GCAACGCAAT  
CCGGCTAAGT AATTACGTG ACCGTGCTGT CCAAAGGCT GACCTTTCGC CCGTCACTCG CGTTGCGTTA

4831 TAATGTGAGT TAGCTCACTC ATTAGGCACC CCAGGCTTGA CACTTTATGC TTCCGGCTCG TATGTTGTGT  
ATTACACTCA ATCGAGTGAG TAATCCGTGG GGTCCGAAAT GTGAATATAG AAGGCCGAGC ATACAACACA

**FIG. 48H**

4901 GGAATTGTGA GCGGATAACA AATTACACACA GGAACACAGCT ATGACCATGA TTACGCCAAG CGCGCAATTA  
CCTTAACACT CGCCTATTGT TAAAGTGTGT CCTTTGTGGA TACTGGTACT AATCGGGTTC GCCTTTAAT

BssHII  
~ ~ ~ ~ ~

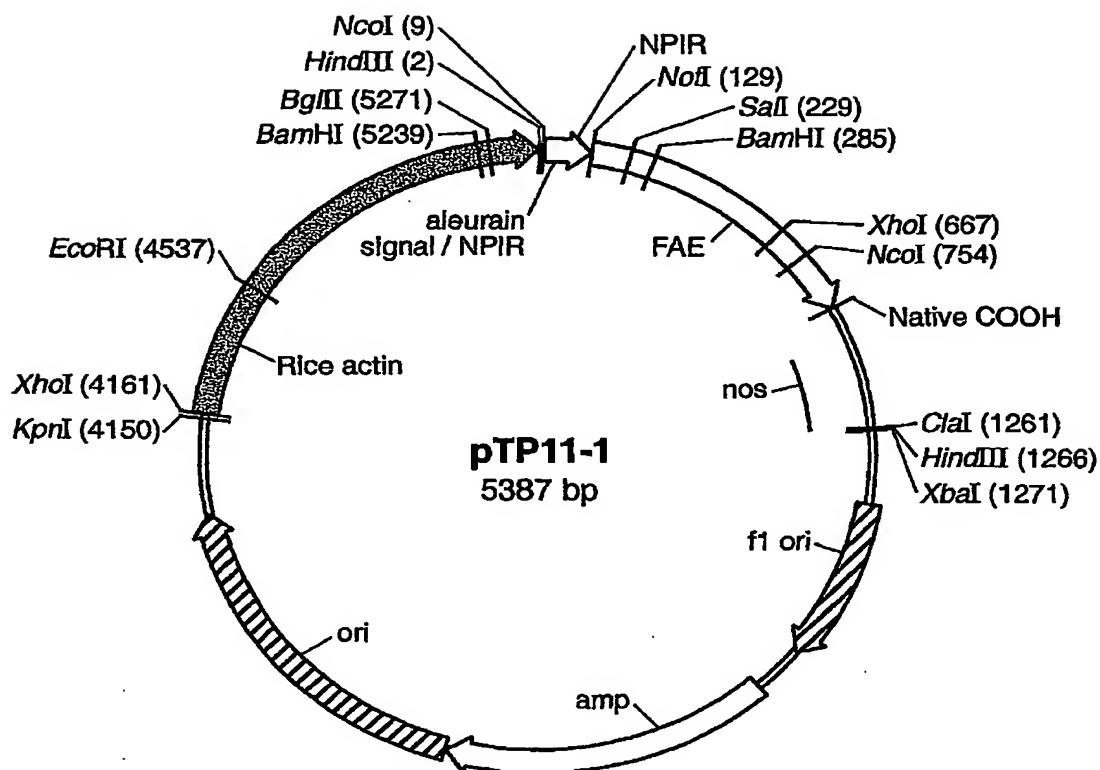
NcoI

4971 ACCCTCAGTA AAGGGAACAA AAGCTGGGTA C  
TGGGAGTGAT TTCCCTTGTT TCGACCCAT G

KpnI  
~ ~ ~ ~ ~

FIG. 48I



**FIG. 49A**

```

NGOI
~~~~~
HindIII
~~~~~
      M A H A R V L L L A L A V L A T A A V A V
1 AAGCTTACCA TGGCCACGCG CCGGCTCTC CTCCTGGCGC TCGCCGTGCT GGCACGGCC GCCGTGCGCG
      NotI
      . A S S S F A D S N P I R P V T D R A A A S T .
71 TCGCCTCCTC CTCCTCCTC GCGACTCA ACCGATCCG GCCGTACCC GACCGCGCG GCGCCTCCAC
      . Q G I S E D L Y S R L V E M A T I S Q A A Y A
141 GCAGGGCATC TCCGAAGACC TCTACAGCCG TTTAGTCGAA ATGGCCACTA TCTCCCAAGC TGCCTACGCC
      Sali
      ~~~~~
 AccI
      ~~~~~
      D L C N I P S T I I K G E K I Y N S Q T D I N G
211 GACCTGTGCA ACATTCCGTC GACTATTATC AAGGAGAGA AAATTACAA TTCTCAACT GACATTACG
      BamHI
      ~~~~~
 . W I L R D D S S K E I I T V F R G T G S D T N .
281 GATGGATCCT CCGCGACGAC AGCAGCAAAG AATAATCAC CGTCCTCCGT GGCCTGGTA GTGATACGAA
 . L Q L D T N Y T L T P F D T L P Q C N G C E V
351 TCTACAATC GATACTAAT ACACCTCAC GCCTTTCGAC ACCCTACAC AATGCAACGG TTGTGAAGTA
 H G G Y Y I G W V S V Q D Q V E S L V K Q Q V S
421 CACGGTGGAT ATTATATTGG ATGGCTCTCC GTCCAGGACC AAGTCGAGTC GCTTGTCAA CAGCAGGTTA
 . Q Y P D Y A L T V T G H X L G A S L A A L T A .
491 GCCAGTATCC GGACTACGCG CTGACCGTGA CCGGCCACAC CCTCGGCGCC TCCCTGGCGG CACTCAGTGC
 . A Q L S A T Y D N I R L Y T F G E P R S G N Q
561 CGCCAGCTG TCTGGGACAT ACGACAACAT CCGCCTGTAC ACCTTCGGCG AACCGCGCAG CGGCAATCAG
 XhoI
      ~~~~~
      A F A S Y M N D A F Q A S S P D T T Q Y F R V T
631 GCCTTCGCGT CGTACATGAA CGATGCCTTC CAAGCCTCGA GCCCAGATAC GACGCAGTAT TTCCGGGTCA

```

FIG.\_49B

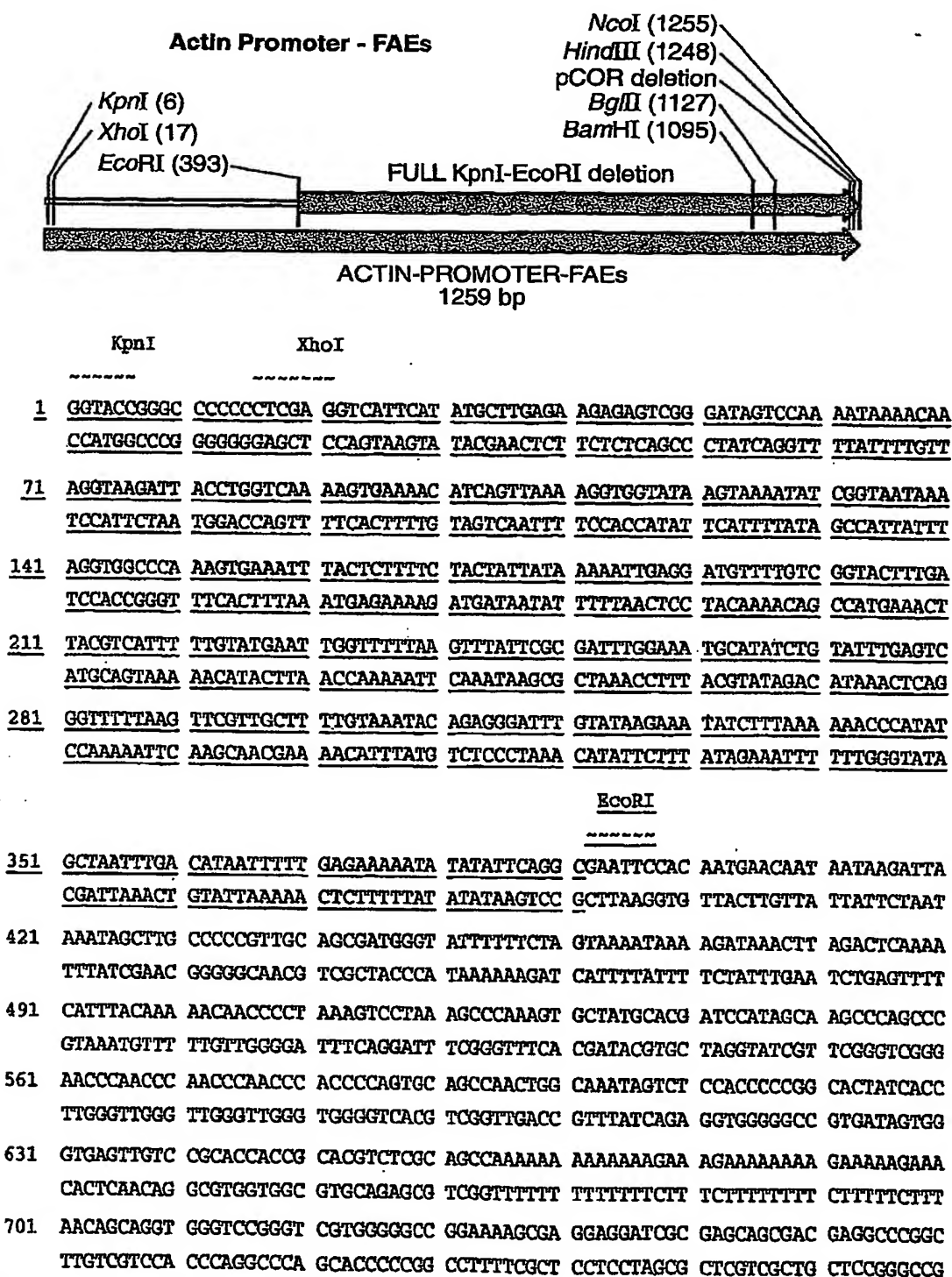
**FIG. 49C**

|       |            |             |             |             |             |            |             |
|-------|------------|-------------|-------------|-------------|-------------|------------|-------------|
| 2451  | TCGATGATGA | CACGTGCGGCC | AACCTACTCTC | TGACAACAGAT | CGGAGGAGCCG | AAGGAGCTAA | CCGCTTTTTT  |
| 2521  | GCACAACATG | GGGGATCATG  | TAACTCGCCT  | TGATCGTTGG  | GAACCGGAGC  | TGAATGAAGC | CATACCAAAC  |
| 2591  | GACGAGCGTG | ACACCACGAT  | GCCGTGAGCA  | ATGGCAACAA  | CGTTGCGCAA  | ACTATTAAC  | GGCGAATAC   |
| 2661  | TTACTCTAGC | TTCCCGGCAA  | CAATTAATAG  | ACTGATGGA   | GGCGGATATA  | GTTGCAGGAC | CACCTCTGCG  |
| 2731  | CTCGGCCCTT | CCGGCTGGCT  | GGTTATATGC  | TGATAAATCT  | GGAGCCGGTG  | AGCGTGGGTC | TCGCGGTATC  |
| 2801  | ATTGCAGCAC | TGGGGCCAGA  | TGGTAAGCCC  | TCCCCTATCG  | TAGTTATCTA  | CACGACGGGG | AGTCAGGCAA  |
| 2871  | CTATGGATGA | ACGAATATGA  | CAGATCGCTG  | AGATAGGTGC  | CTCACTGATT  | AAGCATTTGG | AACTGTCTAGA |
| 2941  | CCAAGTTTAC | TCATATATAC  | TTTAGATTGA  | TTTAAACCTT  | CATTTTAAAT  | TTAAAGGAT  | CTAGGTGAAG  |
| 3011  | ATCCTTTTGG | ATAATCTCAT  | GACCAAAATC  | CCTTAACGTG  | AGTTTTCGTT  | CCACTGAGCG | TCAGACCCCG  |
| 3081  | TAGAAAAGAT | CAAGGATCT   | TCTTGAGATC  | CTTTTTCCT   | GGCGTAATC   | TGCTGCTTGC | AAACAATAAA  |
| 3151  | ACCACCGCTA | CCAGCGGTGG  | TTTGTGTTGC  | GGATCAAGAG  | CTACCAACTC  | TTTTTCCGAA | GGTAACCTGGC |
| 3221  | TTCAGCGAGG | CGCAGATACC  | AAATACTGTC  | CTTCTAGTGT  | AGCCGTAGTT  | AGGCCACCAC | TTCAAGAACT  |
| 3291  | CTGTAGCACC | GCCTACATAC  | CTCGCTCTGC  | TAATCCTGTT  | ACCAGTGGCT  | GCTGCCAGTG | GCATTAAGTC  |
| 3361  | GTGTCTTACC | GGGTGGACT   | CAAGACGATA  | GTTACCGGAT  | AAGCGCAGC   | GGTCGGGCTG | AACGGGGGTT  |
| 3431  | TCGTGCACAC | AGCCAGCTT   | GGAGCGAAG   | ACCTACACCG  | AACGTGATTA  | CCTACAGCGT | GAGCTATGAG  |
| 3501  | AAAGCGCCAC | GCTTCCCGAA  | GGGAGAAAG   | CGGACAGGTA  | TCCGGTAAGC  | GGCAGGGTCC | GAACAGGAGA  |
| 3571  | GCACACGAGG | GAGCTTCCAG  | GGGAAACGC   | CTGGTATCTT  | TATAGTCTTG  | TCGGGTTCG  | CCACCCTCTGA |
| 3641  | CTTGAGCGTC | GATTTTGTG   | ATGCTCGTCA  | GGGGGCGGA   | GCCTATGGAA  | AACGCCCAGC | AACCGGGCCT  |
| 3711  | TTTTACGGTT | CCTGGCCCTT  | TGCTGGCTTT  | TGCTCACAT   | GTTCTTTCTT  | CGCTTATCCC | CTGATTTCTGT |
| 3781  | GGATAACCGT | ATTACCGCCT  | TTGAGTGAGC  | TGATACCGCT  | CGCCGCGACC  | GAACGACCGA | CGCCAGCCGAG |
| 3851  | TCAGTGAGCG | AGGAAGCGGA  | AGAGCGCCCA  | ATACGCAAC   | CGCCTCTCCC  | CGCGCTTGG  | CCGATTCTATT |
| 3921  | AATGCAGCTG | GCACGACAGG  | TTTCCCGACT  | GGAAGCGGG   | CAGTGAGCGC  | AACGCAATTA | ATGTGAGTTA  |
| 3991  | GCTCACTCAT | TAGGCACCC   | AGGCTTTACA  | CTTTATGCTT  | CCGGCTCGTA  | TGTTGTGTGG | AATTGTGAGC  |
| 4061  | GGATAACAAT | TTTCACACAG  | AAACAGCTAT  | GACCATGATT  | ACGCCAAGCG  | CGCAATTAAC | CCTCACATAA  |
| KpnI  |            |             |             |             |             |            |             |
| XhoI  |            |             |             |             |             |            |             |
| 4131  | GGGAACAAA  | GCTGGGTACC  | GGGCCCCCCC  | TCGAGGTGAT  | TCATATGCTT  | GAGAAGAGAG | TCGGGATAGT  |
| 4201  | CCAAAATAAA | ACAAAGGTAA  | GATTACCTGG  | TCAAAAGTGA  | AAACATCAGT  | TAAAAGGTGG | TATAGATATA  |
| 4271  | ATATCGGTAA | TAAAAGGTGG  | CCCAAAGTGA  | AATTTACTCT  | TTTCTACTAT  | TATAAAATTT | GAGGATGTTT  |
| 4341  | TGTCGGTACT | TTGATACGTC  | ATTTTGTGAT  | GAATTTGGTT  | TTAAGTTTAT  | TCGCGATTTG | GAATATGCATA |
| 4411  | TCGTGATTTG | AGTCGGTTTT  | TAAGTTCGTT  | GCTTTTGTAA  | ATACAGAGGG  | ATTGTATATA | GAATATATCTT |
| EcoRI |            |             |             |             |             |            |             |
| 4481  | TAAAAAACCC | ATATGCTAAT  | TTGACATAAT  | TTTTGAGAAA  | AATATATATT  | CAGGCGAATT | CCACAATGAA  |
| 4551  | CAATAATAAG | ATTAATAATG  | CTTGCCCCCG  | TTGCAGCGAT  | GGGTATTTTT  | TCTAGTAAAA | TAAAAGATAA  |
| 4621  | ACTTAGACTC | AAAACATTTA  | CAAAAACAAC  | CCCTAAAGTC  | CTAAAGCCCC  | AAGTGCTATG | CACGATCCAT  |

FIG. 49D

4691 AGCAAGCCCA GCCCAACCCA ACCCAACCCA ACCCAACCCA GTGCAGCCAA CTGGCAATA GTCTCCACCC  
 4761 CCGGCACATAT CACCGTGAGT TGTCCGCACC ACCGCACGTC TCGCAGCCAA AAAAAAAA AGAAAGAAAA  
 4831 AAAAGAAAAA GAAAAACAGC AGGTGGGTCC TCCGCTTCCA AAGAAACGCC CCCCATCGCC ACTATATACA TACCCCCCCC  
 4901 CGACGAGGCC CGGCCCTCCC TCCGCTTCCA CCFACACCA CCACCAACAC CACCTCCTCC CCCCTCGCTG CCGGACGACG  
 4971 TCTCCTCCCA TCCCCCAAC CCFACACCA CCACCAACAC CACCTCCTCC CCCCTCGCTG CCGGACGACG  
 5041 AGCTCCTCCC CCTCCCCCT CCGCCGCCGC CGGTAAACAC CCGCCCTCC TCTCTCTCTT TCTCTCTCTT  
 5111 TTTTCTTCGT CTCGGTCTCG ATCTTTGGCC TTGGTAGTTT GGTGGGCCGA GAGCGGCTTC GTGCCCCAGA  
 BamHI  
 5181 TCGGTGCGCG GGAGGGGCGG GATCTCGCG GATCTCGCTC CCGGCGTGAG TCGGCCCGGA TCCTCGCGGG  
 BglII  
 5251 GAATGGGGCT CTCGGATGTA GATCTCTTTT CTTTCTCTTT TTTGTGGTAG AATTGAATC CCTCAGCATT  
 5321 GTTCATCGGT AGTTTTTCTT TTCAATGATT GTGACAAATG CAGCCTCGTG CCGAGCTTTT TTGTAGC

**FIG. 49E**

**FIG. 50A**

771 CCTCCCTCCG CTTC~~CAAAGA~~ AACGCCCCC ATCGCCACTA TATACATACC CCCCCTCTC CTCCCATCCC  
GGAGGGAGGC GAAGGTTTCT TTGCGGGGG TAGCGGTGAT ATATGTATGG GGGGGGAGAG GAGGGTAGGG

841 CCCAACCCTA CCACCACCAC CACCACCACC TCCTCCCCC TCGCTGCCGG ACGACGAGCT CCTCC~~CCCT~~  
GGGTGGGAT GGTGGTGGTG GTGGTGGTGG AGGAGGGGG AGCGACGGCC TGCTGCTCGA GGAGGGGGGA

911 CCCCCTCCGC GCGCGCCGGT AACCACCCCG CCCCTCTCCT CTTCTTTTCT CCGTTTTTTT TTCTGCTCTG  
GGGGGAGGCG GCGGCGGCCA TTGGTGGGCG GGGGAGAGGA GAAAGAAAGA GGCAAAAAA AAAGCAGAGC

981 GTCTCGATCT TTGGCCTTGG TAGTTTGGGT GGGCGAGAGC GGCTTCGTCG CCCAGATCGG TCGCGGGGAG  
CAGAGCTAGA AACCGGAACC ATCAAACCA CCCGCTCTCG CCGAAGCAGC GGGTCTAGCC ACGCGCCCTC

BamHI

1051 GGGCGGGATC TCGCGGCTGG CGTCTCCGGG CGTGAGTCGG CCGGATCCT CGCGGGGAAT GGGGCTCTCG  
CCCGCCCTAG AGCGCCGACC GCAGAGGCC GCACCTAGCC GGGCCTAGGA GCGCCCTTA CCGGAGAGC

BglII

1121 GATGTAGATC TTCTTTCTTT CTCTTTTGTG TGGTAGAATT TGAATCCCTC AGCATTGTTT ATCGGTAGTT  
CTACATCTAG AAGAAAGAAA GAAGAAAAAC ACCATCTTAA ACTTAGGGAG TCGTAACAAG TAGCCATCAA

HindIII NcoI

1191 TTTCTTTTCA TGATTGTGA CAAATGCAGC CTCGTGCGGA GCTTTTTTGT **AGGTAGAAGC** TTACCATGG  
AAAGAAAAGT ACTAAACACT GTTTACGTCG GAGCACGCCT CGAAAAACA TCCATCTTCG AATGGTACC

KpnI-EcoRI - deletion underlined and restored NCO site in bold in vectors pJQ4.9,  
pJQ3.2 and pJO6.3.

**FIG.\_50B**

**ALEURAIN\_deleted NPIR (Apoplast) Structure and Sequence**

ALEURAIN-NPIR-DEL  
93 bp

+1            M   A   H   A   R   V   L   L   L   A   L   A   V   L   A   T   A   A   V   A  
HindIII NcoI

\*\*\*\*\*

1   AAGCTTACCA TGGCCACGC CCGCTCCTC CTCCTGGCGC TCGCGTGCT GGCCACGGCC GCCGTGCGG  
TTCGAATGGT ACCGGTTCG GCGCAGGAG GAGGACCGG AGCGCACGA CCGTGCCGG CCGCAGCGC

+1 V   A   S   S   R   A   A

NotI

\*\*\*\*\*

71   TCGCCTCCTC CCGCGCGGCC GCC  
AGCGGAGGAG GCGCGCCGG CGG

**FIG.\_51**



## SEE1 ( Senescence enhanced ) PROMOTER sequence

```

1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI          XbaI    XbaI
      ~~~~~
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATG

```

FIG. 52

## SEE1 ( Senescence enhanced ) PROMOTER plus vacuolar aleurain SIGNAL/NPIR sequence

```

1   CATGGGCCAG GTATAATTAT GGGATATCTC AAGCAAATAA TCGAAATATC ACCATTGGCT ACAATATCTG
      PstI          XbaI    XbaI
      ~~~~~
71 AGCTCCGAGT TCTGACTGCA GTCTGGATGA CGCGTGTGT ATCTAGAACT CTAGATAGCA CAGCCACAGC
141 ACCTACAGGA GTGCGACACT TGTGGACTGT AGTAGTGTG GAGACGGAGC TCTTTCCTAC CTCCTGACGT
211 TGCCGCCGTT GTCCATTCCA ACGGCATCAC TCTCAACCAA TCACGCGCTC CCAACAAAAT ATCGTCCCCC
281 ATGTCTTGGC GGAGAGAGAG TACATACATG CTGTGCGGCC GTTTTGTCT GAATCTCGCT TCCACTGGCC
 SmaI
      ~~~~~
351 AATCAGCTCA GCTCCCGGGA GCTCACTCAT TCAAGATCCC ATCGTCGTCG TCACCCCTGG CGTCATGGGA
421 TGGAAAAGAA CCTCCGTTGC TCGGATGAGT CAGCCATATC CCCGAACAGA GTACTGCAAG ATAACCCAAT
      SphI
      ~~~~~
491 TCAGATTCCC CCAATAGAGA AAGTATAGCA TGCTTTCGGG TTTTGTGTTG CTTAATTGAC TTTATTTTGG
561 TTGGAGTTGA ATGCTGATTT GTTGTGTAAA ATGCCCAACC ATCTGAATAT CGAGACGGAT AATAGGCTGG
631 CTAATTAATT TATAGCAAGA TTCTGTAGTG CACATCGCAA ATATCTTTCT GGGCATTACA GCTGGAGGCT
 PstI
      ~~~~~
701 TCATCAGCCT GAAACACTCT GCAGAGCCTG AAGCAAGTGG TGAAGCGTGG CGATGAGATG GGTATAAAAC
771 CCCCGGCACC GGGACGCGAG CTCCCGCCTA CCAGTACCAT CTCGCCTCGC TCCCCCTGCC GGACGACCCA
      M A H G R I L F L A L A V L
841 GTAAAATACT GTTGCCCACT CGCCGGCGAG ATGGCCACG GCCGCATCCT CTTCTTGGCG CTCGCCGTCT
      BssHII
      ~~~~~
 NotI
      ~~~~~
911 . A T A A V A A A S L A D S N P I R P V T E R A .
      TGGCCACCGC CGCGGTGGCC GCCGCATCCT TGGCGGACTC CAACCCGATC CGGCCCGTCA CCGAGCGCGC
      NotI
      ~~~~~
 . A A
981 GGCCGCC

```

FIG. 53

|        |                |
|--------|----------------|
| Intern | Application No |
| PCT/US | 01/43588       |

According to International Patent Classification (IPC) or to both national classification and IPC

Minimum documentation searched (classification system followed by classification symbols)  
IPC 7 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

WPI Data, EPO-Internal, PAJ. BIOSIS

### C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                     | Relevant to claim No. |
|------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X          | US 6 143 543 A (MICHELSEN BIRGIT ET AL)<br>7 November 2000 (2000-11-07)<br>see the whole document                                                                                                                                                                                                                                                                      | 1-3,<br>34-36, 74     |
| Y          |                                                                                                                                                                                                                                                                                                                                                                        | 4-33,<br>37-73        |
| Y          | DE VRIES R P ET AL: "The faeA genes from<br>Aspergillus niger and Aspergillus<br>tubingenensis encode ferulic acid esterases<br>involved in degradation of complex cell<br>wall polysaccharides."<br>APPLIED AND ENVIRONMENTAL MICROBIOLOGY,<br>vol. 63, no. 12, December 1997 (1997-12),<br>pages 4638-4644, XP002203731<br>ISSN: 0099-2240<br>see the whole document | 4-33,<br>37-73        |

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents :

- "A"** document defining the general state of the art which is not considered to be of particular relevance
- "E"** earlier document but published on or after the international filing date
- "L"** document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O"** document referring to an oral disclosure, use, exhibition or other means
- "P"** document published prior to the international filing date but later than the priority date claimed

- **T** later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- **X** document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- **Y** document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- **A** document member of the same patent family

'&' document member of the same patent family

Date of the actual completion of the international search

Date of mailing of the international search report

**27 June 2002**

12/07/2002

Name and mailing address of the ISA  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax (+31-70) 340-3016

Authorized officer

**Grosskopf, R**

## INTERNATIONAL SEARCH REPORT

Inten ☐ Application No  
PCT/US 01/43588

| C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT |                                                                                                                                                                                                                                                                                                         |                       |
|------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| Category *                                           | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                      | Relevant to claim No. |
| A                                                    | <p>GARCIA-CONESA MARIA-TERESA ET AL: "A cinnamoyl esterase from <i>Aspergillus niger</i> can break plant cell wall cross-links without release of free diferulic acids." EUROPEAN JOURNAL OF BIOCHEMISTRY, vol. 266, no. 2, December 1999 (1999-12), pages 644-652, XP002203732<br/>ISSN: 0014-2956</p> |                       |

# INTERNATIONAL SEARCH REPORT

Information on patent family members

Inventor

Application No

PCT/US 01/43588

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s)  | Publication<br>date |
|-------------------------------------------|---------------------|-----------------------------|---------------------|
| US 6143543                                | A                   | 07-11-2000 GB 2301103 A , B | 27-11-1996          |

**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

**BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**